# On the Conservative Influence of Attention on Subjective Perceptual Decision Making

Dobromir Asenov Rahnev

Submitted in partial fulfillment of the

Requirements for the degree of

Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2012

ABSTRACT

On the Conservative Influence of Attention on Subjective Perceptual Decision Making

Dobromir Asenov Rahnev

Current models suggest that perception is a decision process: given noisy perceptual signals, the brain has to decide what they represent. While attention is known to enhance the perceptual signal, it has been unclear how it modulates the decision mechanism itself. Here we explored this issue in a series of studies. We used a spatial cuing paradigm to manipulate the attentional focus of observers, and found that attention leads to a conservative detection criterion such that attended stimuli are reported less often than unattended ones (Chapter 1). We investigated whether this effect would generalize to situations that do not involve detection tasks by using the same cuing paradigm, but instead asking observers to discriminate between two stimulus categories. We found that attention leads to low subjective ratings of visibility (Chapter 2). In both sets of experiments, the results were strongest when detection or discrimination capacity d' was equated between different levels of attention, or when stimuli had low contrast. To account for these results, we developed a variance reduction (VR) model of attention in which attention is postulated to reduce the variability of the perceptual signal, while keeping the decision criteria constant (Chapter 3). The VR model provided a good fit to the data observed in Chapters 1 and 2. We tested critical assumptions of the model using functional magnetic resonance imaging (Chapter 4). We found that high activity in the dorsal attention network (DAN) in the brain, which is indicative of a high attentional state, led to lower variability in the evoked signal in motion sensitive area MT+, thus supporting the idea that attention reduces perceptual variability. Further, high DAN activity resulted in lower confidence ratings, which confirmed that the findings from Chapter 2 generalize to exogenous attentional fluctuations and are not limited to spatial cuing. We

tested the VR model further by extending it beyond the realm of attention (Chapter 5). We used transcranial magnetic stimulation (TMS) to directly increase the variability of the perceptual signal. The effects mirrored the effect of lack of attention: TMS led to decreased performance but increased subjective ratings. Finally, we explored the influence of attention on the amount of information carried by one's subjective ratings. We found that attention made subjective ratings more predictive of accuracy (i.e., attention improved metacognitive sensitivity) despite the fact that it decreased the overall magnitude of the subjective ratings (Chapter 6). To account for this finding, we developed a simple extension to the VR model – the "variance and criterion jitter reduction" (VCJR) model of attention which postulates that attention reduces the amount of trial-to-trial criterion jitter. Computational modeling shows that this reduction of criterion jitter leads to improved metacognitive sensitivity. We discuss these findings in relation to current debates related to attention and subjective perception, and speculate how they may account for our impression that we clearly see everything in our visual fields, including unattended objects that receive little processing.

# Table of Contents

## List of Figures and Tables

## Acknowledgements

I want to thank Hakwan Lau for his incredible dedication in mentoring me. Hakwan made time to talk to me for hours on end when I would randomly drop by his office, explained the ins and outs of signal detection theory, taught me how to analyze fMRI data, and made sure that I received the most credit for my work even if that meant that he received less than he deserved. I cannot imagine a mentor who would be more approachable, more involved, or more supportive of his students. He also had a central role in generating many of the ideas for the studies and models presented here, provided financial support to make data collection and analysis possible, and helped edit the published manuscripts.

I am also grateful to the members of the Lau lab who have helped tremendously with this work. I want to thank Brian Maniscalco for all his help with Matlab over the years, teaching me about computational modeling, and always being willing to discuss obscure issues related to signal detection theory. He also played a truly irreplaceable role in every step of the development of the model described in this manuscript, from the conception and discussion of the basic ideas, to the technical implementation and model fitting; most of the Matlab code used for model comparison and model fitting were also his. Without Brian, this work would not have been possible. I am also grateful to Matthew Davidson and Chrystele Ody for teaching me about neuroimaging, to Guillermo Solovey and Jorge Morales for being such great labmates, and to an army of research assistants including Linda Bahdo, who provided invaluable help with collecting the data for Chapters 4 and 5, as well as Appendices A and B, Tashina Graves, who collected the data for Experiment 2.3, David Zhou, who provided timely help with data collection for the experiment in Chapter 4, Elliott Huang, who ran most of the subjects for the experiments in Chapters 1 and 2, Yoshiaki David Ko, who helped with editing this dissertation, as well as Cynthia Shi, Li Yan McCurdy, Namema Amendi, Cecilia Reyes, Jenessa Holder, and Evan Griffith.

Special thanks are also due to my second lab – that of Floris de Lange in the Netherlands – where I spent three wonderful months and collected some of the data presented here including the studies presented in Chapter 4, and Appendices A and B. Floris has been a truly great and supportive mentor, and has assisted me tremendously with fMRI data collection and analysis. He also provided invaluable help with data collection for Experiment 1.5. Further, Moniek Munneke from the Donders Institute for Brain, Cognition and Behavior made it possible for me to collect the data for Appendix B by providing training and logistic support, while Peter Kok played an invaluable role in the retinotopic mapping used in Appendix B. Finally, I want to thank Bruce Luber for teaching me the intricacies of TMS research, and Sarah Lisanby providing the opportunity for me to conduct the TMS study described in Chapter 5.

I'd also like to thank the members of the Psychology Department at Columbia who have helped make my graduate school experience unforgettable. Bryan Denny, Elliott Wimmer, Jen Silvers, Heather Van Volkinburg, and Karen Kelly have been such a wonderful cohort. Also, special thanks to Patrick Kennedy, Travis Riddle, Diana Keith, Juliet Davidow, Rahia Mashoodh, Jenna Reinen, Caroline Marvin, Erin Kendall Braun, and Jared Van Snellenberg for their support.

Finally, I would like to thank my wonderful girlfriend Sneha Subramanian for all her support and love, and for helping me edit this dissertation. I'd like to also express my appreciation of my parents, Olga and Asen, for being so encouraging every step of the way, and of my brother, Kalin, for always being so protective of me.

## Dedication

To my mom, dad, and brother who are always there for me.

## Preface

Most of the work presented in this dissertation has been taken from published papers and manuscripts in preparation:

Chapters 1-3 are based on findings reported in Rahnev, Maniscalco, Graves, Huang, de Lange, & Lau (2011) Nature Neuroscience.

Chapter 4 is based on Rahnev, Bahdo, de Lange, & Lau (in press) Journal of Neurophysiology.

Chapter 5 is based on Rahnev, Maniscalco, Luber, Lau, & Lisanby (2012) Journal of Neurophysiology.

Chapter 6 is based on Rahnev & Lau (in preparation).

In addition, Rahnev, Lau, & de Lange (2011) Journal of Neuroscience is included as Appendix A, while Rahnev, Kok, Munneke, Bahdo, de Lange, & Lau (in preparation) is included as Appendix B.

Unless otherwise specified in particular cases, I played the primary role in designing and programming the experiments presented here, collected (or supervised the research assistants who collected) and analyzed the empirical data. I also took the leading role in writing and submitting all the manuscripts mentioned above. I discussed the ideas for the studies with a number of people including my advisor, members of the Lau lab, and other researchers. These people also provided invaluable help with generating data analysis ideas, and helping me edit the published manuscripts. I am deeply grateful to all my collaborators that have helped shape this line of research. To acknowledge their contribution in all stages of this research, I have used the pronoun "we" throughout this work.

Some of the data and analyses in Experiments 1.6 and 2.2 were previously presented as a part of my M.A. paper, while some of the behavioral findings from the TMS experiment

(Chapter 5) were reported as a part of work used to satisfy a departmental comprehensive requirement. Reporting these behavioral data and analyses here was necessary in order to introduce the new formal modeling results associated with these studies, as well as some important new analyses on the behavioral data. The introduction of individual data fits and formal model comparisons cast the previously reported data in a new light and, in some cases, gave them different interpretation. Neither of the two appendices was used to satisfy departmental requirements.

# General Introduction

Vision scientists have traditionally shied away from studying the subjective aspects of human perception (Crick and Koch, 2003). One reason for this reluctance may be the role that behaviorism played in the second part of the 20th century. Behaviorism's influence made many researchers reluctant to focus on subjective phenomena in general, and conscious visual perception in particular. Instead, vision scientists focused on characterizing human performance with more accuracy and precision. Methods like signal detection theory (SDT; Macmillan and Creelman, 2005), single- and double-threshold theories (Krantz, 1969), and accumulator models of visual perception (Ratcliff and McKoon, 2008) were developed to explain how humans make perceptual decisions. However, despite the success these theories had in describing and predicting human behavior, they were silent as to the question of the subjective aspect of perception. Many researchers thus tacitly assumed that conscious perception can be fully described by d', drift rate, or similar measures of performance without asking subjects for their subjective impressions of the stimuli.

## Attention and subjective perception

Despite the long-lasting reluctance to study subjective perception, the field has seen a considerable change in direction in the last 30 years. The emergence of cognitive psychology introduced the notion that in order to understand people, one needs to focus on the mind and not only on the outward behavior. This view led to an increasing acceptance that, in addition to performance measures, subjective aspects of perception should also receive the attention of scientists.

Early views on subjective perception equated it with attention (Baars and Laureys, 2005; Frith, 2003; Rees, Kreiman, and Koch, 2002). This idea has been pervasive among both

psychologists and philosophers. Proponents of this view usually point out the fact that withholding attention often results in stimuli failing to enter consciousness, as demonstrated by the phenomena of change and inattentional blindness.

Change blindness (Simons and Levin, 1997) refers to the inability to detect changes that happen in portions of the visual field one is not attendeding to. A large number of studies have demonstrated this effect using a variety of methods. For example, Blackmore, Brelstaff, Nelson, and Trościanko (1995) showed that large changes to a visual stimulus were not detected if they were introduced during saccades. Other studies have used the flicker and the gradual change techniques (Rensink et al., 1997). In the flicker method, two images flash one after the other with a brief blank screen between them. In the gradual change method, the image stays on the screen but a small change is introduced gradually so that it does not draw attention. Surprising demonstrations of the effect have been carried out in real world settings in which subjects often fail to notice that a confederate is substituted with a different person (Levin, Simons, Angelone, and Chabris, 2002).

A related phenomenon is that of inattentional blindness. The term refers to the inability to detect events in plain sight to which one is not paying attention. The most famous demonstration of the effect was reported by Simons and Chabris (1999). Subjects were shown a video of people dressed in black and white T-shirts passing a basketball around, and were instructed to count the number of passes. Despite the conspicuous appearance of a person wearing a full gorilla suit in the middle of the scene, when prompted, about half of the subjects reported completely missing the appearance of the gorilla.

**Dissociation between objective and subjective perception**

Change and inattentional blindness were seen as impressive demonstrations of the intimate relationship between attention and consciousness, and by extension, of the objective

and subjective aspects of perception. Nevertheless, more recently researchers have started to realize that objective and subjective perception can dissociate, sometimes in a startling manner. Some early examples of this came through the discovery of two extreme clinical conditions.

The first of these is named “blindsight” and became a subject of particularly strong interest among vision researchers (Weiskrantz, 1986). Blindsight patients typically have focal damage on their primary visual cortex V1, which makes them blind in the corresponding part of the visual field. However, it was found that these patients were still able to successfully discriminate and respond to stimuli in their blind field of view at a rate significantly better than chance (Weiskrantz, 1986). The disorder quickly became controversial due to the relatively small number of patients with the condition. However, subsequent studies using functional neuroimaging (Leh, Johansen-Berg, and Ptito, 2006), lesions in the monkey primary visual cortex (Schmid et al., 2010), and transcranial magnetic stimulation (Boyer, Harrison, and Ro, 2005) have provided extensive support to the claim that lesions in the primary visual cortex lead to dissociations between objective and subjective components of perception.

The second disorder to highlight subjective experience as a legitimate object of scientific inquiry is essentially the opposite of blindsight. The condition is named “Anton’s blindness” or the “Anton–Babinski syndrome” and describes cases of people who claim to see even when objective tests confirm that they are blind (Roos, Tuite, Below, and Pascuzzi, 1990). This disorder is again caused by lesions to the visual cortex; some theories suggest it emerges when sensory areas become disconnected from language areas (Prigatano and Schacter, 1991).

**Taxonomies of attention and conscious perception**

The discovery of these clinical conditions brought about a wider acceptance of the idea that objective and subjective perception can and do dissociate. One of the results of this acceptance was a re-thinking of the relationship between attention and subjective perception. Several influential researchers now began to argue that attention and consciousness are separate processes that can be dissociated (Crick and Koch, 2003) and are supported by different neural mechanisms (Wyart and Tallon-Baudry, 2008).

In one particularly influential paper, Koch and Tsuchiya (2006) proposed a formal approach to distinguishing between attention and consciousness. The researchers claimed that a double-dissociation exists between the two and that there are processes that require both, neither, or just one of the two. They discussed processes that (1) require both consciousness and attention, (2) require neither consciousness nor attention, (3) require attention but not consciousness, and (4) require consciousness but not attention. We examine these in turn.

Firstly, in relation to processes that require both consciousness and attention, Koch and Tsuchiya gave the example of phenomena like working memory and detection of unexpected and unfamiliar stimuli. In general, full reportability is only possible when both processes are involved.

Secondly, at the opposite end of the spectrum, there are automatic behaviors that do not require or benefit from either attention or consciousness. The lack of necessity for consciousness in such automatic behaviors, like walking or driving a car, has been the reason why they are commonly referred to as “zombie behaviors” (Koch and Tsuchiya, 2006). Further examples of processes that do not require either attention or consciousness are seeing stimuli that are presented too briefly to be attended, and the formation of afterimages.

Thirdly, and more pertinent to the dissociation between attention and consciousness, Koch and Tsuchiya argued that several processes exist that require attention but could operate

in the absence of consciousness. One such process is priming – the facilitation in processing of a stimulus due to exposure to a previous stimulus – which has been shown to require temporal attention but can proceed below the conscious threshold (Naccache, Blandin, and Dehaene, 2002, but see Rahnev, Huang, and Lau, 2012 for evidence that priming could occur in the near absence of spatial attention). Another study to demonstrate a similar effect was conducted by Jiang and colleagues (2006). The researchers presented intact erotic pictures on one side of the visual field, and scrambled pictures on the other side. The presentation of the pictures was masked by continuous flash suppression, so that subjects were not better than chance at identifying which pictures were intact and which were scrambled. However, the observers were better at discriminating the orientation of a Gabor patch presented at the location of the intact compared to the scrambled pictures, suggesting that attention was drawn to the location of the intact erotic pictures without conscious awareness. It has also been shown that the blindsight patient GY has a reaction time advantage for detecting cued targets in his blind field where he has no conscious awareness of the stimuli (Kentridge, Heywood, and Weiskrantz, 2004).

Finally, Koch and Tsuchiya argued that some processes require consciousness but not attention. They found support for this claim in the perception of gist. It is well known that humans extract the gist of a picture (i.e. for example whether it contains animals or not) in as little as 30 ms (Li, VanRullen, Koch, and Perona, 2002) – time insufficient for the deployment of top-down attention (Mack and Rock, 1998). Therefore, Koch and Tsuchiya argued, consciousness, but not attention, is needed to perceive the gist of a scene.

A competing taxonomy was presented by Dehaene, Changeux, Naccache, Sackur, and Sergent (2006). The authors argued that one can distinguish between strong and weak bottom-up stimulus strength as well as whether attentional amplification is present or absent. They based their theory on the widely-used global workspace hypothesis (Dehaene and Naccache,

2001). In its simplest form, the workspace hypothesis states that consciousness arises when a fronto-parietal "workspace" network becomes activated. The workspace neurons have long-distance connections enabling them to broadcast information to many distant areas. In order for the workspace activation to be sufficient for consciousness, it needs to be (1) synchronized: distant areas ought to be oscillating at the same frequency in order to communicate, (2) reciprocal: fronto-parietal regions ought to be able to influence and be influenced by visual areas, and (3) long-lasting: the activation pattern ought to continue for sufficient time to reach a stable state.

According to Dehaene et al. (2006), large stimulus strength, coupled with top-down attentional amplification, gives rise to full-fledged conscious experience. This is the only state that allows for full reportability and is characterized neuronally by long-distance loops and synchronized activity on a global scale.

On the other hand, stimuli that are sufficiently strong but do not receive top-down attentional amplification are theorized to give rise to "preconscious" processing. Similar to the Freudian definition of the term, Dehaene et al. suggested that such processes are not conscious but can quickly enter consciousness if attentional resources become available, even if this happens a few hundred milliseconds after the stimulus presentation. The paradigms of change and inattentional blindness discussed above are good examples of preconscious processing. The events in both paradigms are salient enough that they create robust activations in early sensory areas. According to the researchers, these events create medium-range loops and are stored in buffers for a few hundred milliseconds, but the activation does not propagate to the workspace neurons because of inattention. However, such events can easily enter consciousness if the subject directs his or her attention appropriately.

Lastly, Dehaene et al. argued that subliminal processing is the result of weak or interrupted stimuli regardless of the attentional context. The signal in this case is simply too

weak to trigger a large-scale synchronized state within the workspace neurons. However, this does not imply that attention is completely irrelevant here; rather, Dehaene and colleagues have made it clear that when attention is completely absent, then activations in the brain are extremely meager and basically no priming is possible. However, if attention is present, then activations can reach the semantic level, thus enabling priming to occur.

**Empirical support for the taxonomies of attention and conscious perception**

Despite the differences in Koch and Tsuchiya's and Dehaene et al.'s taxonomies, the central point in both is that attention and consciousness are separate and that in some cases they can act independently from each other.

Strong empirical support for this claim comes from a recent study that manipulated attention and awareness separately in a full factorial 2 x 2 design (van Boxtel, Tsuchiya, and Koch, 2010). Attention was modulated by having observers report on a change of the peripheral stimulus (high-attention condition), or count the number of X's that appeared in a central rapid serial visual presentation task (low-attention condition). Perceptual visibility was modulated by showing, or not showing, a rapidly flickering and rotating high-contrast stimulus in the eye contralateral to the afterimage inducer, while the adaptor was physically present throughout the trial. The results showed that attention and consciousness had opposite effects. In particular, consciously seeing a stimulus (in their case, a grating) increased the afterimage duration, while paying attention to the stimulus decreased the afterimage duration. Thus, this study provided strong evidence that attention and consciousness are distinct processes that can at times even have opposite effects.

**Attention and appearance**

While the debate over whether or not attention and consciousness are separate has been at the forefront recently, a parallel question has increasingly drawn the attention of researchers: given that attention improves object processing, does it also change the subjective appearance of the attended objects?

It has been convincingly established that attention improves detection, discrimination, and identification of degraded stimuli (Bahrami et al., 2008; Blaser, Sperling, and Lu, 1999; Kentridge, Heywood, and Weiskrantz, 2004; Pashler, Johnston, and Ruthruff, 2001; Posner, Snyder, and Davidson, 1980; Reynolds and Heeger, 2009; Wilimzig et al., 2008). This effect has been likened to attention increasing the contrast of the stimuli (Carrasco, Ling, and Read, 2004). However, what is less well established is whether attention changes the subjective appearance of the stimuli.

This question has been debated since the 19$^{th}$ century. William James (1890/1983) and Hermann von Helmholtz (1866) – two of the founding fathers of experimental psychology – argued that attention intensifies sensory impressions. However, Gustav Fechner disagreed and claimed that the opposite was true. Empirical evidence for this debate, however, has only accumulated in the last two decades.

In a seminal paper published in the journal Nature Neuroscience, Carrasco, Ling, and Read (2004) argued that attention indeed changes the appearance of stimuli. The researchers showed a brief exogenous cue at one of two possible locations. Afterwards, two Gabor patches were flashed – one at the cued, and one at an uncued location. Subjects needed to indicate which one of the two stimuli appeared higher in contrast. By systematically varying the contrast of the two stimuli, Carrasco et al. were able to obtain full psychometric functions demonstrating clearly that the cued stimulus was judged to have a higher apparent contrast than the uncued stimulus of the same actual contrast.

There were several aspects of the experiment that further supported the researchers' conclusions. Firstly, subjects were asked to judge the tilt of the Gabor that appeared to have a higher contrast rather than simply indicate which Gabor has a higher contrast. Carrasco and colleagues argued that this manipulation made their subjects less likely to be biased, because they thought that the study was about the perception of tilt rather than about perception of contrast. Secondly, the researchers performed a control experiment in which subjects judged the orientation of the Gabor patch with lower contrast and still found that cued stimuli were judged to have higher contrast. Thirdly, a second control experiment used a much longer period between the exogenous cue and the stimulus (500 ms rather than 53 ms in the main experiment). The results showed that, in this case, the subjects judged the contrast of the cued and uncued stimuli to be equal. These two additional experiments were interpreted as demonstrating that Carrasco et al.'s main finding is not a product of a simple decision-making bias but is a perceptual phenomenon.

Subsequently, the same lab has published multiple papers extending their original finding that attention alters appearance. Using similar procedures, they found that exogenous attention affects the subjective appearance of spatial frequency and gap size (Gobell and Carrasco, 2005), color saturation (Fuller and Carrasco, 2006), motion coherence (Liu, Fuller, and Carrasco, 2006), and flicker rate (Montagna and Carrasco, 2006). Recently, the finding that attention alters appearance has been extended to endogenous attention as well: Liu, Abrams, and Carrasco (2009) used a rapid serial visual presentation (RSVP) to direct subjects' attention to one of two possible locations and again showed that subjects judged the attended stimuli to have higher apparent contrast.

However, other researchers claim that attention does not alter appearance. In particular, Schneider and Komlos (2008) argued that the results of Carrasco et al. were not perceptual but were due to a decision-making bias. According to them, there are four possible ways the

presentation of a cue might affect the perception of contrast: 1. the cue could have a low-level sensory interaction with the stimulus; 2. orienting attention towards the cued stimulus might boost its apparent contrast; 3. attention could also affect the decision mechanism, prioritizing the cued stimulus and causing subjects to tend to report it more frequently as having higher contrast than the uncued stimulus even if the two stimuli were perceptually identical; 4. the presence of a cue differentiates the cued from the uncued stimulus and can also result in a response bias that is independent of attention. Only the second possibility would support the statement that attention alters appearance.

In order to adjudicate among these possibilities, Schneider and Komlos performed an exact replication of Carrasco et al.'s (2004) experiment with the exception that subjects were asked a different question. Rather than indicating which stimulus had a higher contrast, subjects were asked to judge whether the two stimuli appeared to have equal contrast or not. (This type of task is called a same/different task.) This design largely obliterated the decision biases that could be present in Carrasco et al.'s study. Indeed, in the cases in which the two contrasts are perceived to be about equal, subjects in Carraso et al.'s study who are forced to choose one or the other might simply report the cued stimulus as the one with higher contrast because it is more salient. In contrast, subjects in the study by Schneider and Komlos can simply choose to judge the two stimuli as the same, allowing them to choose the "different" option when they are certain.

The results of the study showed that the point of subjective equality (PSE) of estimated contrast coincided with the point of actual equality for both attended and unattended stimuli. In other words, attended and unattended stimuli were now estimated to have the same apparent contrast. Several other experiments also found results that were at odds with Carrasco et al.'s results (Prinzmetal, Nwachuku, Bodanski, Blumenfeld, and Shimizu, 1997; Prinzmetal, Long, and Leonhardt, 2008; Tsal, Shalev, Zakay, and Lubow, 1994). Moreover,

similar to the conclusion reached by Schneider and Komlos, Prinzmetal et al. (1997) suggested that the results from Carrasco's lab may be due to a decision-making bias rather than being due to perceptual processes.

Nevertheless, a recent study argued that the approach by Schneider and Komlos is not as sensitive and may also introduce biases (Anton-Erxleben, Abrams, and Carrasco, in press). The same group also convincingly argued against the conclusions of some of the above papers (see, for example, Carrasco, Fuller, and Ling, 2008). Thus, the debate of whether or not attention changes the appearance of stimuli is ongoing and new developments are likely forthcoming.

**Motivation behind the present research**

We have thus far summarized the state of several debates in the field that relate attention with different kinds of subjective perception. Researchers who work on the taxonomies of attention and consciousness are interested in dissociating the influences of attention and conscious perception, and characterizing what processes depend on either one of these phenomena. Researchers interested in the attentional influence on appearance attempt to characterize how subjects compare strongly and weakly attended stimuli.

Surprisingly, there has been no systematic study of how attention influences the subjective aspects of decision making. The motivation behind the current research is to fill this gap. This issue is important because it has implications about our day-to-day experience. For example, it is known that we have a sense of a uniform visual field, i.e., our visual field appears continuous, and the unattended portions of it do not seem coarser despite the fact that they are not processed as extensively. It is possible that this sense of a uniform visual field is in part due to a bias in subjective perception that related to attention.

**The current questions**

This dissertation explores the topic of the influence of attention on various components of decision making. Here we address this issue by focusing on several specific questions:

(1) Does attention affect the propensity of subjects to report seeing a stimulus? In other words, does attention lead to increased target-present or target-absent responses, or does it have no influence on detection biases? If attention does have an influence, is this influence modulated by the difficulty of the task? These questions are addressed in Chapter 1.

(2) Does attention affect the subjective visibility of the stimulus? Is the attentional influence similar for the objective and subjective aspects of perception? What is the influence of task difficulty? These questions are addressed in Chapter 2.

(3) Can the effects of attention on the propensity to report seeing a stimulus and on subjective ratings be accounted for by a single model? If so, what are the implications for perceptual decision making? These questions are addressed in Chapters 3-5.

(4) Does attention affect the quality of the subjective ratings? One way of measuring this is by examining how well subjective ratings predict accuracy, i.e., by computing subjects' metacognitive sensitivity. So, stated differently, does attention improve metacognitive sensitivity? This question is addressed in Chapter 6.

**The current approach**

Here we present a framework for addressing these questions. We do not attempt to find areas where attention and conscious perception have opposite influences, as did van Boxtel et al. (2010), because this does not necessarily elucidate how attention affects subjective perception. We do not ask subjects to make comparative judgments between attended and unattended stimuli in order to avoid introducing any decisional biases. (The

relation of the current work to these approaches is examined in the General Discussion section.)

Instead, we set to investigate the attentional influence on two separate measures of subjective perception: (1) detection bias and (2) subjective ratings.

Detection bias is best characterized using signal detection theory (SDT; Green and Swets, 1966). SDT treats perception as a decision process of active inference. Given imperfect, noisy perceptual signals, the brain must "decide" what they represent. SDT proposes that such decisions are made based on decision criteria that can be conservative or liberal. For example, if a detection criterion is conservative, a very high activation will be needed to cross the criterion and produce a positive response. On the other hand, if the detection criterion is liberal, a much lower activation can produce a positive response.

Signal detection theory is silent about what factors influence the placement of the decision criteria. Two traditional ways of influencing the decision criteria is through the use of base rate (Ratcliff and McKoon, 2008) or payoff manipulations (Rahnev, Lau, and de Lange, 2011, see Appendix A). The researcher can, for example, inform that in a certain period of the experiment, stimulus A will appear 30% of the time, while stimulus B will appear 70% of the time. If a subject takes this information into account, he or she would move her decision criterion such that more “stimulus B” responses are produced. Similarly, if the subject is paid more for correctly identifying stimulus B than stimulus A (and there is no, or equal, punishment for an incorrectly identifying A or B), then a similar shift of the decision criterion is to be expected.

The above examples demonstrate that criterion setting can be seen as the decisional aspect of perception. However, criterion setting is influenced by many factors beyond cognitive strategy. It has previously been argued that detection criteria may reflect the subjective aspects of perception (Lau, 2007). Indeed, a stimulus that creates a relatively strong

subjective impression may be detected with a liberal criterion, i.e., that stimulus may be reported as present more often than it actually occurs. On the other hand, a stimulus that creates a weaker subjective percept would be detected with a conservative criterion and reported as present less often than it occurs. Further support for this view comes from previous work that has characterized defects in perception, such as blindsight (Weiskrantz, 1986) or neglect (Haligan, Fink, Marshall, and Vallar, 2003), as disturbances of the decision criterion (Gorea and Sagi, 2000).

Despite their potential importance for elucidating subjective perception, decision criteria have traditionally been ignored. They have often been seen as unwanted inter-observer variability that does not reflect performance capacity (Macmillan and Creelman, 2005). In fact, when researchers focus on decision criteria, it is usually in order to construct receiver-operating characteristic (ROC) functions (Ratcliff, McKoon, and Tindall, 1994), i.e., the decision criteria themselves are rarely the focus of primary interest. As a result, little is known about how attention modulates criterion setting, which is central to the perceptual decision mechanism itself (Gold and Shadlen, 2007).

Subjective ratings – the second object of study in this work – have perhaps an even more intuitive and straightforward connection to subjective perception than detection criterion: high subjective ratings are likely indicative of a strong subjective percept and vice versa. Subjective ratings come in several different flavors. The most widespread are confidence ratings and visibility ratings (Merikle, Smilek, and Eastwood, 2001). The former are usually conceptualized as indicating the certainty in one's own decision, while the later as indicating the clarity of the stimulus. However, to obtain subjective ratings, researchers have also used wagering techniques where the subject bets money (or points) on whether or not his or her response was correct (Persaud, McLeod, and Cowey, 2007). In this work, we use the first two types of subjective ratings (confidence and visibility; though see Appendix B for a study

where we used wagering). However, because SDT characterizes all of these ratings in an equivalent way, it is likely that our results generalize to other kinds of subjective ratings as well.

Since the decision criteria will be in the focus of attention in this work, it is important to define here what we mean by the term "decision criterion." We will present two types of studies: detection studies, in which the objective is to determine whether a stimulus appeared or not, and discrimination studies, in which the objective is to determine which of two types of stimuli was presented, and give a subjective rating of certainty. In the former case, "decision criterion" refers to the criterion used for the "target-present" response, and is also referred to as a "detection criterion." In the latter case, "decision criterion" refers to the criterion for differentiating between the two stimulus categories (also referred to as "discrimination criterion"), as well as to the criteria used to differentiate between higher and lower subjective ratings (also referred to as "visibility criteria" or "confidence criteria"). When the term "decision criteria" is used outside of the context of a particular study, then all of the above meanings are intended.

**Outline of the present manuscript**

The present manuscript describes a series of studies investigating how top-down attention influences perceptual decision making. The manuscript is organized into six chapters.

Chapter 1 describes a series of six psychophysical studies investigating the influence of attention on detection criterion. These studies show that attention induces a conservative detection bias wherein subjects tend to report that strongly attended stimuli appear less often than weakly attended ones. This finding is replicated several times under different instructions and stimulus conditions.

Chapter 2 includes three psychophysical experiments that extend the findings from Chapter 1. They demonstrate that attention also leads to low subjective ratings in discrimination tasks. In particular, the studies show that when objective capacity to perform the task is equated, attended stimuli receive lower visibility ratings than unattended ones.

Chapter 3 presents a model that accounts for the findings of Chapters 1 and 2. The model is based on signal detection theory, and postulates that attention decreases the variability of the perceptual signal while keeping the decision criteria constant. The model is therefore named "the variance reduction" (VR) model of attention.

Chapter 4 reports on a functional magnetic resonance imaging (fMRI) study that was designed to test the VR model. The study shows that exogenous fluctuations of attention, as indexed by activity in the dorsal attention network in the brain, lead to similar effects on confidence ratings as the cuing paradigm in Chapter 2.

Chapter 5 extends the VR model beyond attention, and shows that changing the variability of the stimulus directly leads to effects similar to those reported in Chapters 2 and 4. The variability of the signal was increased using transcranial magnetic stimulation (TMS) to the visual cortex.

Chapter 6 introduces a further extension of the VR model to the realm of metacognition. It demonstrates that while attention leads to lower subjective ratings, it improves metacognitive sensitivity as measured by the ability of subjective ratings to predict subjects' accuracy. To account for this finding, the VR model is generalized to include trial-to-trial jitter in the decision criteria that is larger for the unattended stimuli.

The findings from these six chapters are discussed with respect to current theories related to perceptual decision making. Implications pertinent to debates relating attention to consciousness and appearance are also pointed out.

# Chapter 1: Attention Induces a Conservative Detection Bias

## Introduction

It has become increasingly clear that perception is a complex phenomenon. Standard models, such as signal detection theory (SDT; Green and Swets, 1966), treat perception as a decision process of active inference: the brain must decide what the noisy signals represent (e.g., Nikolov, Rahnev, and Lau, 2010). SDT proposes that such decisions are made based on decision criteria that can be conservative or liberal. For example, if a decision criterion in detection (or, "detection criterion") is conservative, a very high activation will be needed to cross the criterion and produce a positive response. On the other hand, if the detection criterion is liberal, a much lower activation can produce a positive response.

Therefore, the detection criterion is of utmost importance to the final outcome of a detection task. Here we set out to investigate how top-down attention influences perceptual decision making by modulating criterion setting. Intuitively, one may think that attention would lead to a more liberal criterion because attention sometimes allows us to notice events that may otherwise be missed. Phenomena such as change and inattentional blindness have further demonstrated observers' tendency to miss very noticeable events when they do not attend to them.

Surprisingly, we found that attention leads to a more conservative bias in detection. In other words, our subjects were more likely to report a stimulus as present when they did not pay attention to it compared to when they did pay attention to it. The effect was particularly strong when one controlled for the overall ability to perform the task, or when contrast was low.

## Methods

### *Participants*

Forty-three subjects in total (22 women; age range = 18–35 years) participated in six psychophysical experiments (Experiments 1.1 through 1.6). Five subjects participated in two of the experiments, while two participated in three of them. All of the subjects were naive regarding the purposes of the experiments, had normal or corrected-to-normal vision, and signed an informed-consent statement approved by the local ethics committee.

### *Materials and Procedure*

All stimuli were presented on a gray background (6.0 cd/$m^2$). In each trial four circles (5° visual angle in diameter) were presented in the four quadrants of the screen with the center of each circle located 5° away from fixation. This configuration was chosen in order to minimize eye movements, as in previous studies (Bahrami, Lavie, and Rees, 2007). A white (27 cd/$m^2$) arrow pointed to one of the two diagonals (Figure 1). Subjects were seated in a dimmed room about 60 cm away from the computer monitor and instructed to maintain fixation at a central red dot (0.4°) displayed on top of the white arrow for the duration of each trial. Stimuli were generated using Psychophysics Toolbox (Brainard,1997) in MATLAB (MathWorks, Natick, MA) and were shown on an iMac monitor (19 inch monitor size, 1680 x 1050 pixel resolution, 60 Hz refresh rate).

The stimuli consisted of a noisy background composed of uniformly distributed intensity values (8% contrast). On top of the noise, we added gratings (0.5 cycles/degree) with probability of 50%. The appearance of the gratings was independent for the cued and uncued diagonals: thus gratings could appear in both diagonals, in neither, or in just one diagonal.

The trials in the main task of the experiment began with 500 ms of a centrally presented pre-cue in the form of an arrow. This pre-cue indicated the likely location of the

relevant stimulus, i.e., the one that subjects were asked to respond to. After 500 ms, four circles were presented for 367 ms (Figure 1; except for Experiment 1.4 where the stimuli were presented for 50 ms). The pre-cued diagonal alternated in blocks of 40 trials. This "blocking" of spatial attention reduced the cognitive demands on the subjects and made it less likely that they could confuse the identity of the cued diagonal. Subjects were asked to indicate whether a grating was present in the diagonal that was highlighted with a response cue. The response-cued diagonal was the same as the pre-cued diagonal on 70% of the trials. Subjects were informed about this fact, which encouraged them to use more attentional resources for the processing of the stimuli in the cued compared to the uncued diagonal.

In Experiments 1.1 to 1.5 the contrast of the stimuli was adjusted online to achieve equal d' for the cued and uncued trials. In Experiment 1.6, the contrast for each subject was fixed throughout the experiment. We used the QUEST staircase procedure (Watson and Pelli, 1983) to find the contrast level for each subject that would produce about 90% correct responses for the cued and uncued stimuli altogether (mean contrast = 2.47%, SD = 0.9%). Each of the 4 runs in this experiment included gratings of fixed contrast – 50%, 67%, 83%, and 100% of the originally obtained contrast value, respectively. The order of the runs was randomized between participants.

Experiments 1.2 and 1.3 included an explicit pay-off structure that encouraged unbiased responding. Subjects were given 1 point for each correct answer (hit or correct rejection) and 0 points for each incorrect answer (false alarm or miss). Experiment 1.3 further included trial-by-trial feedback. In order to further increase subjects' motivation to perform the detection optimally, in each of the two experiments the two subjects with the highest scores were awarded extra $10. In Experiment 1.4 the stimuli were presented for 50 ms to minimize the possibility of eye movements between the stimuli. Finally, in Experiment 1.5 eye movements we measured explicitly using an EyeLink 1000 (SR Research, Osgoode, ON,

Canada) infrared camera recording at 1200 Hz. The eye-tracker had gaze resolution of 0.01° (noise limited) and gaze position accuracy of 0.5°.

In each of Experiments 1.1 to 1.6, subjects completed 960 experimental trials separated into 4 runs of 6 blocks. Feedback was given at the end of each block consisting of 40 trials, except for Experiment 1.3 where feedback was given after each trial.

Our task was relatively demanding and our subjects were untrained in psychophysical tasks. Overall eight subjects needed to be excluded because of inability to perform better than chance (1, 0, 2, 1, 0, and 4 subjects were excluded from Experiments 1.1 to 1.6, respectively).

*Statistics*

For most analyses we used a two-sided paired-sample t-test. For the psychophysical experiments we computed the signal detection measures d' and criterion c. We first coded each trial into a hit, miss, correct rejection, or a false alarm. Then, we computed the hit rate (HR; equals #hits / [#hits + #misses]) and false alarm rate (FAR; equals #false alarms / [#false alarms + #correct rejections]) (Macmillan and Creelman, 2005). Then d' was computed as:

$$d' = z(HR) - z(FAR)$$

and criterion c was computed as:

$$c = -0.5 * (z(HR) + z(FAR))$$

where z is the inverse of the cumulative standard normal distribution that transforms HR and FAR into z scores.

## Results

*Experiment 1.1: Effect of attention on the detection criterion*

We used a spatial cuing paradigm (Figure 1) to investigate the effects of attention on decision criterion in a detection task. Previous studies on attentional cuing mostly focused on detection sensitivity, and when they investigated decision criterion, the results can be somewhat difficult to interpret because sensitivity can influence criterion, as has been demonstrated empirically (Dusoir, 1983) even though the two are mathematically independent (Macmillan and Creelman, 2005). Therefore, a change in criterion under attentional cuing could be indirectly due to its effect on sensitivity. To isolate these effects, in Experiment 1.1 to 1.5 the contrast of the stimuli was titrated in such a way that sensitivity (d') was equated between the cued and uncued conditions. To achieve this, the cued locations contained lower contrast stimuli (see Methods).

We first confirmed that our subjects were attending preferentially to the cued diagonal by investigating whether they required a lower stimulus contrast for cued stimuli to achieve the same level of performance as for the uncued stimuli. The results showed that for each of the nine subjects the average contrast for the cued stimuli (mean contrast = 5.5%) was lower than the contrast for the uncued stimuli (mean contrast = 7.8%) with the difference being significant for the group ($p = 0.02$). Further, we confirmed that there was no significant difference in discriminability (d') for the cued (d' = 1.6, SD = .17) and the uncued (d' = 1.7, SD = 0.45) stimuli ($p > 0.1$).

In the critical test, we compared our subjects' criteria for claiming that they had seen a grating in the attended and unattended trials. We used the classic signal detection theory measure for criterion called c (Green and Swets, 1966). A negative c means that subjects are liberal, i.e. they claim that they have seen the stimulus more often than it actually appears,

while a positive c means that subjects are conservative. A c of zero demonstrates that the participant is unbiased. Our subjects had a positive c for the cued trials (c = 0.4) and a negative c for the uncued trials (c = -0.34) with the difference being significant in the group ($t(7) = 2.89$, $p = 0.023$; Figure 2). Thus, it appears that when objective information processing in the brain is controlled for, weakly attended stimuli are detected with lower criterion than strongly attended ones.

One may wonder if the effects could be influenced by the presence of grating targets in the locations to which subjects were not probed to respond, and these gratings might have acted as distractors (Palmer and Moore, 2009; Baldassi, Megna and Burr, 2006). To address this question, we analyzed separately the trials in which the "unprobed" location did or did not contain the grating targets (Figure 3). We found that for both types of trials, subjects had both higher hit and false alarm rates for the uncued stimuli. Therefore, our results above cannot be due to the influence of the unprobed stimuli.

*Experiments 1.2 and 1.3: Providing explicit payoff*

Experiment 1.1 demonstrated that subjects have different response biases for detecting strongly and weakly attended stimuli. However, subjects were never actually encouraged to be optimal (i.e., unbiased), so it is hard to determine whether these results were due to perception or decision biases. We tested whether the same results would be obtained when explicit pay-off structure was defined.

In Experiment 1.2 we gave subjects one point for every correct answer and subtracted one point for every wrong answer; everything else was same as Experiment 1.1. This pay-off structure was chosen to encourage unbiased responding. Indeed, it can be mathematically shown that in this situation, subjects' expected score is maximized by an unbiased response

strategy (i.e., by setting c = 0; Macmillan and Creelman, 2005). We explained this to our subjects and encouraged them to be as unbiased as possible.

The results, however, showed the exact same pattern as Experiment 1.1. Subjects had a positive c for the cued trials (c = 0.32) and a negative c for the uncued trials (c = -0.3; Figure 4a). Critically, the difference in criterion between cued and uncued trials was statistically significant ($t(5) = 3.1$, $p = 0.027$), while there was no difference in d' ($p > 0.1$).

In Experiments 1.1 and 1.2, subjects were given feedback after each block of 40 trials, but not after each individual trial. In Experiment 1.3 we used the same design as in Experiment 1.2, but gave subjects feedback after each trial. Nevertheless, subjects remained conservative in cued trials (c = 0.26) and liberal in uncued trials (c = -0.21), with the difference between these trials being still significant ($t(9) = 3.08$, $p = 0.013$; Figure 4b). Again, d' was matched across the two types of trials ($p > 0.1$).

Further, we checked whether subjects were able to become more unbiased over the course of the experiment. We analyzed the first half and second halves of the experiment separately and found that there was almost no change in detection criteria: criterion for the cued trials changed from c = 0.289 to 0.263, while for the uncued trials it changed from c = -0.196 to -0.252. A direct comparison of the bias in the first and second half of Experiment 1.3 showed that subjects' bias to use a more liberal criterion for the uncued trials did not change between the first and second parts of the experiment ($p = 0.8$). Thus, it appears that the divergent biases exhibited in attended and unattended trials are robust and cannot be overcome even with trial-by-trial feedback.

*Experiments 1.4 and 1.5: Effects not due to eye-movements*

In Experiments 1.1 through 1.3 we presented the stimuli for 367 ms. Even though we instructed subjects to fixate in the center of the screen, and used a stimulus presentation

design that encouraged the maintenance of proper fixation, it is still possible that subjects broke fixation during the stimulus presentation period. Therefore, in Experiment 1.4 we used a presentation time of 50 ms (other details being similar to Experiment 1), so that subjects could not break fixation as easily. We again achieved similar results: subjects remained conservative in cued trials (c = 0.61) and liberal in uncued trials (c = -0.13), with the difference between these trials being significant ($t(4) = 3.65$, $p = 0.022$; Figure 4c). Again, d' was matched across the two types of trials ($p > 0.1$).

Finally, in Experiment 1.5 we investigated the issue of eye-movements more directly and actually monitored subjects gaze throughout the experiment. We found that subjects made an eye-movement of at least one degree (the center of the stimuli was 5 degrees away from fixation) on only 1% of the trials. After excluding these trials, we still found the same results: positive criterion for cued trials (c = 0.24), negative criterion for uncued trials (c = -0.19) and a significant difference between the two ($t(5) = 3.45$, $p = 0.018$; Figure 4d) with lack of significant difference in d' ($p > 0.1$).

Further analysis of the eye-tracking data showed that the average standard deviations across all subjects were 0.23 and 0.55 for the horizontal and vertical dimensions, respectively (see Figure 5). To ensure that eye-movements did not contribute to our results on criterion, we correlated across subjects the propensity to make eye movements with the magnitude of the criterion difference between cued and uncued trials. It turned out that the correlation was not significant and it was, if anything, in the opposite direction such that less eye movement was associated with bigger criterion difference ($r = –0.36$, $p = 0.49$). Thus, eye movements could not account for the effects of attention on criterion.

*Experiment 1.6: Addressing the contrast confound*

An important confound in Experiments 1.1 through 1.5 was that in order to equate capacity d', we used different contrast levels for cued and uncued stimuli. Thus, it is possible that the differences in criterion c were due to the differences in contrast rather than the attentional level. To investigate this issue we conducted Experiment 1.6 in which we used the same multiple contrasts for both the cued and uncued locations. As an initial test of our data, we examined the effect of attention on detection capacity (d'). We compared the cued and uncued trials for each contrast in each subject. Not surprisingly, detection capacity was much higher in the cued compared to the uncued trials ($t(35) = 6.01$, $p < 0.0001$), confirming that attention increases capacity d'.

We then considered the influence of attention on the detection criterion when effects of signal enhancement were controlled for. To this end, we investigated the detection criterion for each level of detection capacity. We plotted detection capacity against detection criterion for each level of contrast (Figure 6) and investigated the lines of best fit for each subject. As can be seen from the figure, the linear fits suggest that higher attention resulted in a higher (more conservative) criterion for capacity levels of up to d' of 3.7 (~97% correct). The y-intercepts for the fits for cued and uncued stimuli were significantly different for d' of up to 2 (~85% correct; $p$'s $< 0.05$). For higher values the relationship inverses but it should be noted that this is purely due to the nature of the linear fit since the highest observed d' for the uncued stimuli was 3.01 - lower than the point at which the relationship inverses.

Further, it appeared that attention may have changed the relationship between detection criterion and detection capacity. Indeed, the line of best fit had a significantly negative slope for the attended trials ($t(8) = 3.26$, $p = 0.01$) but not for the unattended trials ($t(8) = 0.43$, $p = 0.43$), and the difference in slopes between these two types of trials was

marginally significant ($t(8) = 2.07$, $p = 0.07$). Thus, the relationship between d' and detection criterion was found to depend on the attentional level of the subject.

As further evidence that the results above were not exclusively due to a stimulus confound, we next investigated the effect of attention on the detection criterion for each level of contrast. We compared the detection rates for cued and uncued trials, excluding four conditions which produced extreme differences in detectability between the cued and the uncued stimuli (d' difference was bigger than 2). Intriguingly, a paired-sample t-test demonstrated that our subjects were much more liberal in the uncued trials ($t(31) = 5.18$, $p < 0.0001$). The percent hits ($p = 0.05$) and percent false alarms ($p < 0.0001$) were also higher for the uncued trials. Thus, it appears that subjects tended to report uncued stimuli as present much more often than cued ones, even when contrast level was not a confound.

We then tested whether the above effects were primarily driven by stimuli of certain contrast level. We plotted the criterion c for each contrast separately (Figure 7). As can be seen from the figure, cued stimuli were detected more conservatively when they had low contrast. Specifically, for the lowest two contrasts the difference in criterion was significant (both $p$'s $< 0.01$), while for the highest two levels of contrast it was not (both $p$'s $> 0.1$). The difference disappeared for high-contrast stimuli presumably because the task was easy enough that subjects could do it almost perfectly and without significant bias.

These results show that attention modulated the relationship between contrast and criterion. As above, this was further confirmed by the fact that the line of best fit had significantly negative slope for the cued trials ($t(8) = 4.39$, $p = 0.002$) while the slope was not different from zero for uncued trials ($t(8) = 1.57$, $p > 0.1$). A direct comparison of the two slopes demonstrated that the slope for the cued trials was significantly smaller than the slope for the uncued trials ($t(8) = 3.11$, $p = 0.015$). This shows that the effect of contrast was not the

same for the cued and uncued stimuli, and therefore, the results of Experiments 1.1 through 1.5 could not be explained merely by the difference in contrast level.

To understand further the influence of attention on criterion, we examined hit and false alarm rates separately (Figure 8). A multiple regression with factors of attention, contrast, and subject-specific effects demonstrated that contrast level modulated hit rates ($p < 0.001$) but not false alarm rates ($p = 0.48$). In other words, higher contrast did not lead to fewer false alarms though it did lead to more hits. We return to this finding in Chapter 3.

Finally, we performed a multiple regression analysis to look at the effect of attention on criterion when other factors were regressed out. It turned out that when d', contrast, and subject-specific effects were controlled for (i.e., regressed away as effects of no interest), higher attention was still significantly predictive of a more conservative criterion ($p < 0.0001$).

**Discussion**

Results from six experiments demonstrated that attention modulates perceptual decision making in a surprising fashion. While attention increases one's capacity d', it also changes the decisional aspect of perception, thus leading to lower detection rates. In Experiment 1.1 we showed that subjects were more liberal in detecting uncued stimuli: they reported these stimuli as appearing more often that cued ones. Experiment 1.2 demonstrated that these results remain even when subjects are given explicit payoff that encourages them to be unbiased, while Experiment 1.3 showed that not even trial-by-trial feedback was able to eliminate this suboptimal bias. Experiment 1.4 demonstrated that these effects generalized to variations of stimulus presentation times and Experiment 1.5 confirmed that our results were not due to eye-movements. Finally, Experiment 1.6 showed that these results were not due to the influence of contrast and confirmed that attention leads to a conservative detection criterion even when other factors such as contrast and d' were controlled for.

What remains unclear, however, is whether this conservative bias is specific to detection tasks, or whether a parallel effect could be found in different kinds of tasks. In particular, if attention acts to decrease the subjective experience of the subject, then one may expect that in discrimination tasks attention would result in lower subjective ratings. We tested for this possibility in Chapter 2.

# Chapter 2: Attention Leads to Decreased Subjective Ratings

## Introduction

Experiments 1.1 through 1.6 demonstrated that subjects have a tendency to report weakly attended uncued stimuli as present more often that strongly attended cued ones. One explanation for these findings is that they reflect perceptual rather than cognitive processes. In particular, it is possible that under certain conditions weakly attended stimuli produce higher subjective visibility than their strongly attended counterparts. We tested for this possibility in a more direct way in two studies in which we asked subjects to give subjective ratings of visibility to cued and uncued stimuli. A third study examined this question without the use of attentional cuing; instead attention was manipulated by extending it to more or fewer objects. To anticipate, we found that, as predicted, attention led to lower subjective ratings, thus confirming that the results from Chapter 1 were based on perceptual rather than cognitive processes.

## Methods

### *Participants*

Fifty-two subjects in total (32 women; age range = 18–40 years) participated in three psychophysical experiments (Experiments 2.1-2.3). All subjects were naive regarding the purposes of the experiments, had normal or corrected-to-normal vision, and signed an informed-consent statement approved by the local ethics committee

### *Materials and Procedure*

Experiment 2.1 was similar to Experiment 1.1, while Experiment 2.2 was similar to Experiment 1.6. The main difference was that subjects were asked to indicate whether the tilt

of the gratings (which were always presented) was 45° or 135°, and then indicate the visibility of the tilt of the gratings (Figure 9). In Experiment 2.1 we used a 2-point scale (high/low), while in Experiment 2.2 we used a 4-point scale (1 – not visible at all; 4 – highly visible). We were careful in explaining to the subjects that they should rate the visibility of the tilt of the grating rather than the overall brightness of the stimulus. In Experiment 2.1 the contrast of the cued and uncued stimuli was updated online as in Experiments 1.1 through 1.5. In Experiment 2.2 we chose fixed levels of contrast but unlike Experiment 1.2, for simplicity the contrasts here were chosen to be the same for all participants (1.7, 2.2, 2.7, and 3.2% contrast) and were not separated in different runs.

As in Experiments 1.1-1.6, subjects completed 960 experimental trials separated into 4 runs of 6 blocks. Feedback was given at the end of each block consisting of 40 trials. A total of four subjects needed to be excluded because of inability to perform better than chance (3 subjects in Experiment 2.1 and 1 subject in Experiment 2.2).

In Experiment 2.3 the stimuli were presented for 33 ms. After a delay of 500 ms, subjects saw a response cue that instructed them on which stimulus they should do the task. Subjects had to indicate the tilt (left/right) of a Gabor patch and rate the visibility (high/low) of the tilt of that patch (Figure 13). Subjects completed 8 blocks of 125 trials each for a total of 1000 trials. Within each block there were always either 2 or 4 patches. We computed d' and stimulus visibility for each of the 5 levels of contrast that we used in this experiment.

Our task was relatively demanding (especially in Experiment 2.3) and our subjects were untrained in psychophysical tasks. Overall 16 subjects needed to be excluded because of inability to perform better than chance (3, 1, and 13 subjects were excluded from Experiments 2.1 to 2.3, respectively). Including these subjects did not change the results of the experiments.

*Statistics*

As in Experiments 1.1-1.6, most tests that we performed were paired-sample t-tests. In order to compute the signal detection measure d' we again first coded each trial into a hit, miss, correct rejection, or a false alarm. Trials in which subjects reported that the stimulus was right tilted were coded as hits if the grating was indeed right tilted and as false alarms if the grating was left tilted. Trials, in which subjects reported that the stimulus was left tilted, were coded as misses if the grating was right tilted and as correct rejections if the grating was left tilted. We then computed hit and false alarm rates, as well as d', as in Experiments 1.1-1.6.

**Results**

*Experiment 2.1: Effect of attention on visibility ratings*

In Experiment 2.1 subjects experienced the same presentation sequence as in Experiments 1.1-1.5 with two differences (Figure 9). First, the gratings were presented on every trial and had either left or right tilt; subjects' task was to discriminate between these two possibilities. We used a diagonal design so that the cued stimuli were presented in one diagonal, while the uncued stimuli were presented in the other diagonal. The orientation of the gratings was chosen randomly for each diagonal. Second, after indicating their best guess for the direction of the tilt, participants indicated the visibility of the stimuli ("high visibility" or "low visibility").

Subjects' performance in the cued trials was slightly higher than in the uncued trials, but the difference was not significant (d' difference = 0.05, $p = 0.27$). Critically, despite their slightly better performance for cued trials, subjects rated the visibility of the cued trials significantly lower than the visibility of the uncued trials (difference = 0.25, $t(8) = 3.0$, $p = 0.017$; Figure 10). Thus, it appears that the weakly attended stimuli indeed produce higher subjective visibility.

*Experiment 2.2: Effect of attention on visibility ratings as a function of contrast*

As before, Experiment 2.1 contained a stimulus confound: the contrast for the uncued stimuli was higher than for the cued stimuli. That design was chosen in order to equate the capacity d' between the two conditions. Therefore, in Experiment 2.2, we used several contrast levels, that were the same for the cued and uncued trials, in order to investigate the influence of contrast and attention at the same time.

As in Experiment 1.6, we found that stimulus discriminability (d') was much higher in the cued compared to the uncued trials ($t(23) = 9.9$, $p < 0.0001$), confirming that attention increases capacity d'.

We then considered the influence of attention on the visibility ratings. We plotted detection capacity against visibility ratings for each level of contrast (Figure 11) and investigated the lines of best fit for each subject. As can be seen from the figure, the linear fits suggest that higher attention resulted in lower visibility ratings for capacity levels of up to d' of 2.7 (~91% correct). The evidence is clearest for d' values between 0.5 and 1.5 since for lower and higher values there is virtually no overlap between the d' values for cued and uncued stimuli.

Similarly to Experiment 1.6, we confirmed that these results were not exclusively due to a stimulus confound. We performed a multiple regression in order to isolate the effect of attention on visibility ratings. We included attention, d', contrast, and subject-specific effects as predictors. When the effects of the other factors were regressed away, higher attention was predictive of *lower* visibility ratings ($p < 0.005$).

As further evidence that the results above were not exclusively due to a stimulus confound, we examined the visibility ratings for each level of contrast (Figure 12). Surprisingly, for the two lowest contrasts, visibility was judged to be higher for the uncued

than the cued stimuli ($t(5) = 3.72$, $p < 0.01$). This relationship reversed as the contrast of the gratings increased and for the highest contrast the cued stimuli were judged significantly more visible ($t(5) = 2.8$, $p < 0.05$). These results mirror the results that we obtained in Experiment 1.6. Namely, for low contrasts subjects are more liberal for the uncued stimuli and they also judge these stimuli as more visible. For stimuli of higher contrast, subjects become unbiased in the detection task, and the cued stimuli are now judged as more visible.

Finally, as in Experiment 1.6, we confirmed that attention modulated the relationship between contrast and visibility ratings. Indeed, the slope of the line of best fit in Figure 12 for the cued trials was significantly bigger than that for the uncued trials ($t(5) = 3.37$, $p = 0.02$). In other words, the visibility of the gratings increased for both the cued and the uncued gratings with the increase of contrast but this increase was significantly bigger for the cued stimuli.

*Experiment 2.3: Effects of attention do not depend on cuing*

Experiments 1.1-1.6 and 2.1-2.2 manipulated attention by explicitly cuing the relevant location. Therefore, it is possible that the effects that we observed depend on presence of an attentional cued.

To test this, in Experiment 2.3 we presented different number of stimuli on the screen in order to manipulate how subjects distribute their attention to different objects (Figure 13). In one condition we used 2 items on the screen (a relatively focused mode of attention), and in the other we used 4 items on the screen (a relatively distributed mode of attention). Both d' and visibility ratings increased with higher contrast (Figure 14). Further, d' was roughly similar for the 2-patch task with 6% contrast, and the 4-patch task with 8% ($p = 0.95$), 10% ($p = 0.7$), and 12% ($p = 0.99$) contrast (see the horizontal dashed line in Figure 13). Nevertheless, compared to the 6%-contrast 2-patch task, the visibility of the grating was judged to be higher for the 4-patch task for the 8% contrast ($p = 0.01$), 10% contrast ($p = 0.01$), and 12% contrast

($p = 0.02$). Thus, similar to Experiment 2.1, less attention (in the 4-patch task) led to higher subjective stimulus visibility ratings even though discrimination capacity (d') was matched.

**Discussion**

The experiments reported here extended the results from Experiments 1.1-1.6 and demonstated that uncued stimuli receive higher visibility ratings. Experiment 2.1 showed that when objective capacity d' is equated for cued and uncued stimuli, the uncued stimuli receive significantly higher subjective ratings of visibility. Experiment 2.2 demonstrated that this effect is not due to the different contrasts used, and that instead, attention modulated the relationship between contrast and subjective ratings. Finally, Experiment 2.3 showed that the results of the previous two experiments are not due to the specifics of the cuing paradigm that we used, and that similar results can be obtained by spreading attention more or less widely (and thus having more or less attentional resources for each stimulus).

These results were predicted based on the findings from Chapter 1. Further, in both sets of experiments, the results were strongest when capacity d' was equated between different levels of attention, or when stimuli had low contrast. Therefore, Experiments 2.1-2.3 help establish a link between criterion setting for detection and subjective experience and suggest that the same underlying phenomenon could be causing both of these effects.

However, what remains unclear is the nature of the mechanism through which attention leads to conservative subjective biases. In Chapter 3 we present a model that attempts to provide such a mechanism.

# Chapter 3: Variance Reduction (VR) Model of Attention

## Introduction

The psychophysical studies from Chapters 1 and 2 demonstrated that attention leads to a conservative detection criterion, coupled with deflated subjective ratings. These findings are surprising as they are not predicted by popular theories of attention (e.g., Desimone and Duncan, 1995). In order to explain these results, we designed a signal detection theoretic model. The two critical assumptions of the model are that subjects use unified criteria for both the cued and uncued locations and that attention reduces the trial-by-trial variability of the internal perceptual signal (Figure 15). This decrease in variance enhances the signal-to-noise ratio but also reduces the occurrence of high signal trials caused by chance fluctuations. When the same unified criterion is used for target detection in both conditions, the decreased variance for cued trials makes them less likely to cross the detection criterion, which entails fewer target-present responses in detection experiments and fewer high confidence responses in discrimination experiments. We call our model "the variance reduction" (VR) model of attention to highlight that one of its central assumptions is that attention reduces perceptual variability.

## Computational modeling assumptions

We modeled the results from Experiments 1.1, 1.6, 2.1, and 2.2 with a computational model based on signal detection theory (SDT). We made several standard assumptions: (1) the two stimuli used in the experiment gave rise to internal signals normally distributed along some decision axis; (2) perceptual decisions were made by comparing the signal on some decision axis to a criterion; (3) confidence judgments were made by comparing the signal on some decision axis to multiple criteria, corresponding to the multiple confidence ratings

available to subjects in this experiment; and (4) criteria for perceptual decisions and confidence ratings were set in the same way for cued and uncued stimuli. The last assumption derives from previous research (Gorea and Sagi, 2000) which has demonstrated that subjects tend to use a single set of criteria for different sets of stimuli even if they are clearly labeled and spatially separated.

**Model specifications**

In our SDT model attention modulated both the signal and the noise of the internal representations. Thus, in this model attention changed both the distance (μ) between the Gaussian distributions and the standard deviation (σ) of the distributions (hence, we refer to this model as 'mean+SD'). The standard deviation for the uncued stimuli ($\sigma_{uncued}$) was always set to 1. When the model was applied to Experiment 1.1, it included four free parameters: $\sigma_{cued}$, $\mu_{uncued}$, a parameter that quantified the increase of $\mu_{uncued}$ with attention, and the location of the detection criterion. When applied to Experiment 1.6, it included seven free parameters: $\sigma_{cued}$, $\mu_{uncued\_contrast}$ for each of the four levels of contrast (4 parameters), a parameter that quantified the increase of each $\mu_{uncued}$ with attention, and the location of the detection criterion. Applied to Experiment 2.1, the model contained six free parameters: $\sigma_{cued}$, $\mu_{uncued}$, a parameter that quantified the increase of $\mu_{uncued}$ with attention, and the criteria levels used for discrimination and visibility judgments (3 parameters). Finally, when applied to Experiment 2.2, the model contained thirteen free parameters: $\sigma_{cued}$, $\mu_{uncued\_contrast}$ for each of the four levels of contrast (4 parameters), a parameter that quantified the increase of each $\mu_{uncued}$ with attention, and the location of the criteria levels used for discrimination and visibility judgments (7 parameters corresponding to the 8 possible answers; the 8 answers are produced by combining the 2 stimulus choices and 4 visibility levels).

To test if our model only produced good fits because of the number of free parameters, we compared it to two simpler SDT models. In the first one, attention was allowed to affect only the distance (μ) between the Gaussian distributions ('mean-only' model), while in the second attention was not allowed to modulate anything ('null' model). In all four applications, the 'mean-only' model had one fewer parameter than the 'mean+SD' model, which allowed both the mean and SD of the distributions to be modulated by attention: it lacked the $\sigma_{cued}$ parameter. The 'null' model had two fewer parameters than the 'mean+SD' model: it lacked both the $\sigma_{cued}$ and the parameter that quantified the increase of $\mu_{uncued}$. All other parameters were identical across the 3 models. The point of this comparison was to show that the extra parameters in the 'mean+SD' were necessary and worth the extra complexity.

Following standard SDT methods, we assumed that the 2-dimensional stimulus image was reduced to a single scalar value representing the likelihood that this stimulus was a target or a non-target. We did not explicitly model this information reduction. Our SDT model operated at the level of overall probabilities of giving each possible response following the presentation of each kind of stimulus and did not address the trial-by-trial pattern of responses.

**Model fitting**

We fit the models to the data using a maximum likelihood estimation approach that has previously been used within a signal detection framework (Dorfman and Alf, 1969). Briefly, the likelihood of a set of signal detection model parameters given the observed data can be calculated using the multinomial model. Formally,

$$L(\theta|data) \propto \prod_{i,j} Prob_{\theta}(Resp_i|Stim_j)^{n_{data}(Resp_i|Stim_j)}$$

where each $Resp_i$ is a behavioral response a subject may produce on a given trial, and each $Stim_j$ is a type of stimulus that may be shown on that trial.

The expression "$n_{data}$ ($Resp_i$ | $Stim_j$)" is a count of how many times a subject actually produced $Resp_i$ after being shown $Stim_j$.

The expression "$Prob_\theta$ ($Resp_i$ | $Stim_j$)" denotes the probability with which the subject produces the response $Resp_i$ after being presented with $Stim_j$, according to the signal detection model specified with parameters θ. According to SDT, in an experiment with two possible stimuli and n levels of confidence subjects have 2*n possible responses and therefore set 2*n-1 decision criteria that allow them to determine how to categorize each new trial. We denote these monotonically increasing criteria as $c_{1,} c_{2,} \ldots c_{2*n-1}$. In addition let $c_0 = -\infty$, and $c_{2*n} = \infty$. The ordering of the set of response types "Resp" follows the ordering of response types defined by setting the monotonically increasing set of criteria $c_{1,} \ldots c_{2*n-1}$ on a decision axis. For instance, in Experiments 1 and 2, $Resp_1$ = "target absent" and $Resp_2$ = "target present"; and in Experiment 3, $Resp_1$ = "high visibility, left tilt", $Resp_2$ = "low visibility, left tilt", $Resp_3$ = "low visibility, right tilt", $Resp_4$ = "high visibility, right tilt." Then, if we assume that $Stim_j$ gives rise to a Gaussian distribution with a mean $\mu_j$ and standard deviation $\sigma_j$, the expression $Prob_\theta$ ($Resp_i$ | $Stim_j$) evaluates to:

$$Prob_\theta(Resp_i|Stim_j) = \int_{c_{i-1}}^{c_i} \frac{1}{\sqrt{2\pi\sigma_j^2}}\, e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}}\, \mathrm{d}x$$

Note that the models were not fit to summary statistics of performance such as percent correct or average visibility. Rather, models were fit to the full distribution of probabilities of each response type contingent on each stimulus type. Various kinds of summary statistics (e.g. d', c, percent correct, average visibility ratings, and so on) can be derived from this full behavioral profile of stimulus-contingent response probabilities.

We fit all models under consideration to the observed data by finding the maximum-likelihood parameter values θ. Maximum likelihood fits were found using a simulated annealing procedure (Kirkpatrick, Gelatt and Vecchi, 1983). Model fitting was conducted separately for each subject's data. The estimation procedure was reliable; subsequent repetitions of the model fitting procedure produced negligible variations in the parameter estimates for each model of each subject's data.

**Formal model comparison**

The maximum likelihood associated with each model characterizes how well that model captures patterns in the empirical data. However, comparing models directly in terms of likelihood can be misleading; greater model complexity can allow for tighter fits to the data but can also lead to undesirable levels of overfitting, i.e., the erroneous modeling of random variation in the data. Therefore, we compared the models using the Bayesian Information Criterion (BIC). This measure provides a means for comparing models on the basis of their maximum likelihood fits to the data while correcting for model complexity (Burnham and Anderson, 2002).

**Modeling results**

We constructed a signal detection theoretic model in which attention increased the magnitude of the perceptual signal and decreased its trial-by-trial variability (see above for details). We fit the model to the data for each observer and plotted the mean predicted value across observers for each of Experiments 1.1, 1.6, 2.1, and 2.2. The model was able to capture the pattern of the data in all experiments (Figure 16). In Figure 16b,d we marked the contrasts with the numbers 1 to 4 where 1 stands for the lowest presented contrast and 4 stands for the

highest contrast. Based on the model fits across the four experiments, attention seemed to reduce the variability (standard deviation) of the cued distributions by a factor of ~2.

We tested the intuition that the critical feature of the model was this reduction in noise. We investigated if models in which attention does not affect the variability of the signal were able to explain the data. The results showed that the model in which attention modulated both the mean and SD of the distributions ('mean+SD' model) outperformed models in which attention only modulated the mean of the distributions ('mean-only' model) or had no effect ('null' model). This was true even when we used model selection methods (Bayesian information criterion; BIC) to punish models for complexity. Thus our model provided good fits not just because it has more parameters. The averaged posterior probabilities (estimated from BIC) for the 'mean+SD' model were 0.91, 0.77, 0.56, and 0.93 for Experiments 1.1, 1.6, 2.1, and 2.2, respectively. These values were higher than the values for the 'mean-only' (0.04, 0.06, 0.20, and 0.07) or the 'null' (0.05, 0.17, 0.24, and 0.00) models, providing strong evidence that the attentional modulation of the signal variability is indeed the critical feature of our model.

**An intuitive account for the VR model's ability to fit the results from Chapters 1 and 2**

A depiction of the signal detection theoretic model that accounted for our findings is provided in Figure 17. The VR model postulates that attention leads to smaller trial-by-trial variability in the internal perceptual signal. Here we try to give an intuitive account of how the VR model is able to explain the results of the in Chapter 1 and 2.

*Explaining the data from Experiments 1.1-1.5*

In Experiments 1.1-1.5 we equated the detection capacity (d') of the cued and uncued stimuli by presenting higher contrast stimuli in the uncued locations. We found that subjects

produced higher hit and false alarm rates for the uncued stimuli (see Figure 3). In Figure 17a the higher contrast of the uncued stimuli is represented by a bigger distance between the Target-Absent (red; left) and Target-Present (blue; right) distributions. This larger difference between the cued and uncued trials is counterbalanced by the larger variability of the distributions for the uncued trials, so that d' (signal-to-noise ratio) for cued and uncued trials is equated. The VR model also assumes a single unified criterion used for both cued and uncued trials (previous studies have shown that subjects indeed tend to use a single unified criterion for detection of targets in two possible locations, compromising optimality for both; see Gorea and Sagi, 2000). As depicted in Figure 17a, these assumptions lead to conservative detection for the cued stimuli and liberal detection for the uncued stimuli, which is what the experimental data showed. The optimal (unbiased) criteria for both conditions are depicted with dashed lines.

*Explaining the data from Experiment 1.6*

In Experiment 1.6, contrast was kept the same for cued and uncued stimuli (resulting in a higher d' value for the cued stimuli). Subjects were more conservative in detecting the cued stimuli, with the effect being largest for low stimulus contrasts. Unlike Figure 17a where the distance between the peaks of the distributions was larger for uncued stimuli (because of the difference in contrast between cued and uncued stimuli), in Figure 17b one would expect this distance to be higher for cued stimuli because attention leads to boosting of the perceptual signal (and contrast was the same for cued and uncued stimuli in Experiment 1.6). Our results showed that subjects tended to avoid excessive false alarms by fixing their unified detection criterion with respect to the Target-Absent distribution to achieve a relatively constant level of false alarms (Figure 8; see Green and Swets, 1966). Because the Target-Absent distribution for uncued trials is wider, this leads to more false alarms for uncued trials (see shaded

regions). However, the hit rate depends on the contrast presented (Figure 8). Specifically, for low contrast levels the Target-Present distributions for both cued and uncued stimuli are relatively close to the Target-Absent distributions and the unified criterion falls to the left of the peak of the Target-Present distribution. In this case, due to the higher variability of the Target-Present distribution for uncued stimuli, hit rate would be higher for uncued stimuli. However, for higher contrast levels the Target-Present distribution's peak would be to the right of the unified detection criterion which would lead to lower hit rate for uncued stimuli. In Figure 17b the unified criterion falls on the left of the peak of the Target-Present distributions which leads to higher hit rate for cued stimuli (i.e., a condition similar to using high contrast).

*Explaining the data from Experiment 2.1*

In Experiment 2.1, we again equated the discrimination sensitivity d' and observed higher visibility ratings for uncued stimuli. According to signal detection theory, a discrimination task can be represented with Gaussian distributions as in Figure 17a,b. In this case, the Target-Present and Target-Absent distributions would correspond to the distributions for right- and left-tilted patches. To give visibility (or confidence) ratings in a discrimination task, subjects can set additional "visibility" criteria next to the unified criterion for discrimination (see also Figure 15). High visibility ratings will be given for any trial that produces an internal response that falls to the right of the rightmost visibility criterion or to the left of the leftmost visibility criterion.

If the relative lack of attention leads to larger variability of internal perceptual signal for the uncued stimuli, the wider spread means that more trials will exceed the visibility criteria for high visibility ratings (Figure 17c). Since the two visibility criteria behave similarly, in the following discussion we focus on the rightmost visibility criterion. For that

criterion, high confidence trials coming from the Target-Present distribution correspond to hits in the detection task depicted above, while high confidence trials coming from the Target-Absent distribution correspond to false alarms in the detection task. Therefore, in the model the uncued stimuli would produce higher visibility ratings for the same reasons for which uncued stimuli receive more hit and false alarms in Experiments 1.1-1.5.

*Explaining the data from Experiment 2.2*

Finally, in Experiment 2.2 we used the same contrasts for the cued and uncued stimuli. As in Experiment 1.6, this resulted in a higher d' value for the cued stimuli. Also, similarly to Figure 17b, in this experiment one would expect the distance between the two Gaussian distributions to be higher for cued stimuli because attention boosts the perceptual signal (Figure 17d). Finally, according to our model, as in Experiment 1.6 subjects would give more high visibility ratings (graphically corresponding to hit and false alarms in the detection experiment in Figure 17b) for the uncued stimuli when the contrast is low but not when the contrast is high, which is again what we found experimentally.

**Models in which attention does not affect the variability of the distributions cannot account for the observed data**

As discussed above, the 'mean-only' model did not provide good model fits. Here we give an intuition as to why that model could not fit the observed data. We focus on the detection experiments, though similar argument could be given for the discrimination experiments.

Figure 18a provides a depiction of the 'mean-only' model. The 'mean-only' model allowed for attention to increase the mean of the Target-Present distribution, without changing the mean of the Target-Absent distribution. The assumption that attention does not alter the

mean of the Target-Absent distribution is based on the underlying neuronal physiology. Indeed, when a non-target is presented in the uncued condition, we expect the firing to be near baseline so that attention would not have much room to further decrease the firing. Given the above assumptions, the 'mean-only' model predicts that cued and uncued trials will have similar levels of false alarms, and that cued trials should have a higher hit rate than uncued trials. However, Experiments 1.1 and 1.6 showed a very different pattern of results. Specifically, the false alarm rate was higher for uncued trials (see Figures 3 and 8). Hit rates were also higher for the uncued trials in both Experiment 1.1 and the first two contrast levels in Experiment 1.2, which is the opposite of what the 'mean-only' model would predict.

Figure 18b provides a depiction of an alternative version of the 'mean-only' model which can mathematically account for the observed results. This alternative model allows attention to move both the Target-Absent and Target-Present distributions in either direction. In order for this model to account for the findings, in Experiment 1.1 attention would need to shift the Target-Absent distribution 0.7 standard deviations to the left. (It is impossible to analyze the influence of attention on the Target-Present distribution because the contrast differed between the cued and uncued stimuli.) In Experiment 1.6, attention would need to shift the Target-Absent distribution 0.6 standard deviations to the left. The Target-Present distribution would need to move 0.4 standard deviations to the left for the lowest two contrasts, and 0.03 and 0.4 to the right for the highest two contrasts, respectively. Thus, in order for the alternative 'mean-only' model to fit the data, the overall effect of attention would need to be to decrease the means of the Target-Absent and Target-Present distributions. This effect would correspond to attention having a largely suppressive influence on neuronal firing. However, such an effect does not fit with the well documented findings that attention boosts the firing rate. Even for the highest contrast in Experiment 1.6, the boost in neuronal firing for the Target-Present distribution would be smaller than the suppression for the Target-

Absent distribution. Further, this kind of alternative 'mean-only' model would not be able to account for our results from the discrimination studies, since a move of both the distributions to the left (or to the right) would result in a bias for choosing either left- or right-tilted gratings and such a bias was not observed in the data. Therefore, because of its neural implausibility and lack of generalizability to all our data, we did not implement this alternative 'mean-only' model computationally.

**Conclusion**

The variance reduction (VR) model presented here accounts for our counter-intuitive results from Chapters 1 and 2. We have presented formal model fitting and model comparison, as well as an intuitive account for the ability of the model to fit the data. In the following two chapters we test the model with two paradigms that are very different from the design of the experiments in Chapters 1 and 2. We find that the VR model generalizes to a wide variety of experimental designs, and even in situations that do not involve attentional manipulation.

# Chapter 4: Extending the VR Model to Spontaneous Fluctuations of Attention

## Introduction

The data from Chapter 2 show that attention can lead to low subjective visibility ratings. In Chapter 3 we proposed a formal model – the variance reduction (VR) model – that accounts for this effect. According to the model, lack of attention increases the variability of the internal perceptual response and confidence ratings are given using the same unified criteria for both attended and unattended stimuli. Thus, the "low attention" distributions are marked by high variability. This makes these distributions, as compared to the "high attention" distributions, extend further into the high confidence regions and thus producing more high confidence trials.

Here we used fMRI to investigate the neural bases of the above effects. In particular, we tested whether spontaneous fluctuations of attention would lead to differential effects on accuracy and confidence ratings, as well as whether such effects would be accompanied by a decrease in variability of the perceptual signal (as predicted by the VR model). We measured the pre-stimulus BOLD activity in the dorsal attention network (Corbetta and Shulman, 2002) which is implicated in directed attention and working memory and is part of the task-positive system in the brain. The name "dorsal attention network" distinguishes it from the "ventral attention network" which is specialized for the detection of behaviorally relevant stimuli, particularly when they are salient or unexpected (Corbetta and Shulman, 2002). The fluctuations of BOLD activity in dorsal attention network have been investigated in a number of previous resting-state studies (Fox et al., 2005, 2006, 2007; Sadaghiani et al., 2009, 2010) and have been theorized to reflect the attentional level of the subject (Eichele et al., 2008; Sapir et al., 2005). As in previous studies (Hesselmann, Kell, Eger, and Kleinschmidt, 2008; Hesselmann, Kell, and Kleinschmidt, 2008; Sadaghiani et al., 2009), we focused on pre-

stimulus activity in the dorsal attention network because it is not contaminated by stimulus related activity and reflects ongoing neural fluctuations.

Based on our previous findings, we hypothesized that high pre-stimulus BOLD activity in the dorsal attention network would be associated with low confidence ratings and higher accuracy. Further, we predicted that these effects would be accompanied by an increase in the variability of the perceptual signal as measured by the variance in the evoked responses in the motion sensitive region MT+.

## Methods

### *Participants*

Fifteen volunteers (9 women; mean age = 22; range = 19-26 years) participated in the experiment. All subjects were naive regarding the purposes of the experiments, had normal or corrected-to-normal vision, and signed an informed-consent statement approved by the local ethics committee (CMO region Arnhem-Nijmegen, The Netherlands).

### *Stimulus and Procedure*

In each trial (Figure 19), subjects were required to indicate the overall direction of motion of white dots (density = 2.4 dots/degree$^2$; speed = 6 degrees/sec) presented inside a black annulus (outer circle radius = 10°; inner circle radius = 1°). The motion direction was either contracting or expanding. Incoherent dots moved randomly with the same speed as coherent dots. All dots had infinite lifetime: they were never removed from the screen during the motion presentation. A small fixation square was presented for the duration of the trial and subjects were required to maintain fixation on it. The stimuli were presented on gray background, and were generated using the Psychophysics Toolbox (Brainard, 1997) in MATLAB.

We employed a sparse event-related design (Hesselmann, Kell, Eger, and Kleinschmidt, 2008; Hesselmann, Kell, and Kleinschmidt, 2008; Sadaghiani et al., 2009) where each trial began with 20-40 seconds of no motion stimuli. This period allowed for the blood-oxygenated level dependent (BOLD) signal to return to baseline after each trial. We then presented 100 ms of coherent motion and followed it with a "mask" of 400 ms of random motion. The masking was employed to increase task difficulty and to avoid retrospection based on iconic memory (Sperling, 1960). We elected to use motion stimuli because of ease of control of the stimulus properties and localization of the relevant brain regions. These have also been used in previous studies of spontaneous fluctuations of BOLD activity (Hesselmann, Kell, and Kleinschmidt, 2008; Sapir et al., 2005). Subjects were required to indicate the direction of motion (expanding / contracting), and give a confidence rating with separate button presses made with their right hand. Confidence was judged on a 4-point scale, where 1 represented low and 4 represented high confidence. Subjects were instructed to use the scale as evenly as possible. Nevertheless, since some subjects had a bias toward using one side of the confidence scale, and in order to maximize power, we defined what constituted high and low confidence for each subject individually using a median split. Overall low confidence was defined as a rating of 1 or 2 for 9 subjects and as a rating of 1 for the other 6 subjects. If both button presses were not completed within 6 seconds, the trial was marked as unanswered and excluded from further analyses. Subjects completed 104 trials separated in two scans of 52 trials. Each scan took about 26 minutes. After these two scans, we acquired a 307-volume "resting state" scan (Fox and Raichle, 2007) that took about 10 minutes.

Prior to the fMRI experiment, each subject took part in a one-hour training session on a separate day, in which the subject practiced 520 trials that were identical to the trials in the scanner but for a shorter ITI of 1-3 seconds. During the first half of the training session subjects were given trial-by-trial feedback. The feedback was discontinued during the second

part of the training in order to prepare subjects for the experiment in the scanner environment. Based on the data from the training session, motion coherence levels were chosen for each subject to produce ~65% correct responses (mean motion coherence = 30.6%, SD = 27.5%). Just before the fMRI experiment, subjects practiced for additional 5 minutes (68 trials) to remind them of the task.

*Behavioral analyses*

We analyzed our data to check for the existence of trial-to-trial dependencies. For each subject we investigated whether a correct or high confidence response on the current trial predicted correct or high confidence response on the subsequent trials. To do that, we estimated the proportion of correct and high confidence trials following correct, error, high confidence, and low confidence trials. We then compared these proportions for correct and error trials, as well as for high and low confidence trials. The comparison was done between subjects using a paired-sample t-test.

*fMRI acquisition*

Images were acquired on a 3 Tesla Trio MRI system (Siemens, Erlangen, Germany). Functional images were acquired using a 32-channel coil, with a single shot gradient echo-planar imaging (EPI) sequence (repetition time: 1950 ms; echo time: 30 ms; 31 ascending slices; voxel size: 3 x 3 x 3 mm; flip angle = 80 degrees; field of view = 192 mm). A high resolution anatomical image was acquired using a T1-weighted MP-RAGE sequence (repetition time: 2300 ms; echo time: 3.03 ms; voxel size: 1 x 1 x 1 mm).

*fMRI preprocessing*

Analysis was performed using SPM5 (Wellcome Department of Imaging Neuroscience, London, UK). The first 6 volumes of each scan were discarded to allow for scanner equilibration. Preprocessing consisted of realignment through rigid-body registration to correct for head motion, slice timing correction to the onset of the first slice, co-registration of the functional and anatomical images, segmentation of the anatomical image, normalization to Montreal Neurological Institute (MNI) space using the grey matter image obtained from the segmentation, interpolation of functional images to 2 x 2 x 2 mm, and smoothing with a Gaussian kernel with a full-width at half-maximum of 8 mm.

*Definition of brain networks*

We defined the attention, the default, and the alertness networks using a separate 10-minute resting state scan and employing seed-based analyses as in previous studies (Fox et al., 2005, 2006; Sadaghiani et al., 2009, 2010). Briefly, using the MarsBaR toolbox for MATLAB, we extracted the time courses for the resting state scan for spheres with 10 mm radius. We used seeds in the right intraparietal sulcus (IPS; 27, -58, 49) for the attention network (Fox et al., 2006; Sadaghiani et al., 2009), the posterior cingulate cortex (PCC; -5, -49, 40) for the default network (Fox et al., 2005; Sadaghiani et al., 2009), and dorsal ACC (dACC; 0, 15, 40) for the alertness network (Sadaghiani et al., 2010). Each time course was high-pass filtered (1/128 Hz), and the time courses for all grey matter voxels, all white matter voxels, and all cerebral spinal fluid (CSF) voxels were regressed out in a multiple regression (Sadaghiani et al., 2009). This step was similar to the "global signal regression" that is typically performed in such experiments (Fox et al., 2005) whereby the average signal from the whole brain is regressed out. This pre-processing step is useful in that it removes BOLD activity from physiological (i.e., non-neuronal) origin (Fox et al., 2009) but it has been criticized for

inducing artificial anti-correlations between networks (Anderson et al., 2011; Murphy et al., 2009). Nevertheless, the primary interest in this study was investigating the influence on behavior of the pre-stimulus activity in a specific network (the dorsal attention network) and thus we wanted to isolate its influence independent of physiological and other global influences. On the other hand, we were not concerned with potential anticorrelations that could result from this processing step.

The resulting time courses were used as regressors in separate general linear models (GLM) for each subject; the three networks were defined as the set of voxels correlated at $p < 0.001$ with the respective regressor. Each of the three networks was then combined at the second level. Each subject-specific network was a combination of that subject's network masked with the corresponding group network.

To test for the robustness of the findings, the dorsal attention network was alternatively identified using independent components analysis (ICA). We employed the GIFT toolbox (Calhoun et al., 2001) for MATLAB to extract the networks consistent across all subjects (Damoiseaux et al., 2006). We used the Infomax algorithm to find 16 components. The attention network was identified among these 16 components and was defined using a z-score threshold of 1. The region MT+ was removed from the attention networks defined either using either seed-based methods or ICA.

*Pre-stimulus activity analyses*

After the networks were defined for each subject, we analyzed the data from the first two runs in which subjects engaged in the motion discrimination task. We extracted time courses for each network using the MarsBaR toolbox (http://marsbar.sourceforge.net/) and regressed out nuisance variables such as head motion and global signal fluctuations as above. Pre-stimulus activity was defined as the average of the two volumes preceding the onset of the

motion stimulus. We used paired sample t-tests and repeated measures ANOVAs to test whether pre-stimulus activity differed significantly between high and low confidence trials, as well as between correct and incorrect trials. Even though we tested for the effects in three different networks (see above), we did not correct for multiple comparisons because our main hypothesis was about the effects in the dorsal attention network, and the other networks were just included for completeness.

We also investigated whether pre-stimulus activity in the dorsal attention system influenced the variability in MT+. For each subject, we categorized trials as either having low or high pre-stim dorsal attention network activity if the pre-stimulus activity in the network was lower or higher than the average activity for that subject. We then computed the Fano factor of the MT+ evoked activity for these two types of trials. The evoked activity was defined as the maximum BOLD activity of the 3 scans that had highest activity on average (across all subjects and all trials). The Fano factor was defined as the variance divided by the mean for the evoked activity for each of the two types of trials for each subject (Carandini, 2004; Churchland et al., 2010; Gur et al., 1997).

*MT+ localizer*

We identified MT+ using a separate localizer scan after the main experiment and the resting state scan. Sixty blocks of moving dots (block duration of 16 s) were alternated with 10 blocks of stationary dots (block duration of 16 s), resulting in ~19 min scan duration. Motion-sensitive areas were obtained using the contrast moving dots > stationary dots. Local maxima near the ascending limb of the inferior temporal sulcus were defined bilaterally and combined to form area MT+ separately for each subject (Hesselmann, Kell, and Kleinschmidt, 2008). MT+ defined in this was used as a region of interest (ROI) in a separate GLM analysis (see below).

*Standard GLM analysis*

Regressors for the first-level analysis of evoked activity were obtained by convolving the unit impulse time series for each condition with the canonical hemodynamic response function. The motion stimulus was modeled with duration of 500 ms. Our model included 4 regressors, reflecting the combination of 2 levels of accuracy (correct/error), and 2 levels of confidence (high/low). We included 12 nuisance regressors related to head motion: three regressors related to translation and three regressors related to rotation of the head, as well as their derivatives (Lund et al., 2005).

*Simulations*

We performed simple computer simulations in order to investigate whether the VR model can provide an adequate explanation of the results, especially given that we did not observe a statistically significant effect of pre-stimulus BOLD signal in the dorsal attention network on motion discrimination accuracy (see Results). The main idea of the VR model is that attention decreases the variability of the perceptual signal. We implemented this intuition in the equations below. We formalized that:

$$S(t) = E(t) + N_P(t, A(t))$$

where S is the evidence available to the perceptual system, E is the evidence present in the stimulus, $N_P$ is the amount of physiological noise inherent in the system, A is the level of attention in the current trial, and t indicates the trial. The pre-stimulus BOLD signal in the dorsal attention network was modeled as:

$$B(t) = e_A * A(t) + N_B(t)$$

where B is the BOLD signal, $N_B$ is the noise present in the BOLD signal, and $e_A$ controls the effect of attention on the BOLD signal. We modeled the level of attention present in each trial (A) using a uniform distribution in the interval (0, 1). The evidence present on each trial (E) was sampled from a normal distribution with a mean $\mu_E$ and standard deviation $\sigma_E$ in order to reflect the fluctuations in the presentation of the random dot motion. $N_P$ and $N_B$ were modeled as Gaussian distributions with means $\mu_P$ and $\mu_B$, and standard deviations $\sigma_P$ and $\sigma_B$, respectively. Since the absolute amount of evidence was not of interest, $\mu_E$ was set to 1 and $\mu_P$ was set to 0. For simplicity we assumed a linear influence of attention on the decrease of variability in the perceptual signal and therefore modeled $\sigma_P$ as $\sigma_P = m - d*A$, where, m is the mean value of the noise in $N_P$ while d controls the degree of attentional decrease of the variability in $N_P$. Finally, since the random variables E and $N_P$ are both normally distributed, their sum will also be normally distributed and have a standard deviation that is the sum of the standard deviations of E and $N_P$. Therefore, rather than fitting $\sigma_P$ and $\sigma_E$ separately, we fit only their sum.

The perceptual decision was made by comparing S(t) to 0: positive values indicated correct responses, while non-positive values were coded as incorrect response. On the other hand, confidence was determined by comparing S(t) to a set of criteria ($c_1$, $c_2$). High confidence ratings were given if $S(t) < c_1$ or if $S(t) > c_2$. Because of lack of significant bias for expanding or contracting motion in our experiment, for simplicity $c_1$ was set to $-c_2$. Thus, the set of confidence criteria we used was (–c, c).

We fitted the above equations with the average data from all 15 participants using a simulated annealing procedure (Kirkpatrick et al., 1983). Each iteration of the fitting procedure generated 1,500,000 trials (roughly corresponding to 1000 times the data from our experiment) based on the current values of the parameters. The fitting function attempted to minimize the error in overall percent of correct answers, overall confidence, the mean pre-

stimulus BOLD activity for all high confidence, low confidence, correct, and error trials. To test for the robustness of our findings, the fitting was done four different times using different starting values for the parameters. We varied the starting values of the parameters in order to avoid getting "stuck" in local minima. Each of the four fits provided a good fit to the data. Here we report the results from the first fit which produced the best match to the observed values; the results of the other fits were consistent with the best one. Using the parameters from the best fit, we generated 10,000 experiments that consisted of 15 subjects each completing 100 trials (15,000,000 trials total) and checked the power to detect significant effects on confidence and accuracy.

**Results**

We first checked for the existence of trial-to-trial dependencies in accuracy or confidence ratings. Paired-sample t-tests demonstrated that accuracy on the current trial predicted neither accuracy ($t(14) = -0.83$, $p = 0.42$) nor confidence ($t(14) = -1.3$, $p = 0.21$) on the subsequent trial. Similarly, confidence on the current trial predicted neither accuracy ($t(14) = 0.2$, $p = 0.85$) nor confidence ($t(14) = -1.04$, $p = 0.32$) on the subsequent trial. Thus, it appeared that there were no reliable trial-to-trial dependencies indicating that our subjects were generating stochastic behavioral responses.

We identified the dorsal attention network using standard seed-based methods (Figure 20, see Methods; other methods of identifying the network, such as independent component analysis, ICA, gave similar results, see Table 1). Confirming our prediction, we found that high pre-stimulus BOLD activity in the dorsal attention network was associated with low confidence ratings ($F(1,13) = 6.68$, $p = 0.02$; Figures 21 and 22). On the other hand, pre-stimulus activity in this network was not reliably associated with accuracy ($F(1,13) = 0.45$, $p = 0.52$). There was also no interaction between confidence and accuracy ($F(1,13) = 0.02$, $p =$

0.89). The above analyses were performed using repeated measures ANOVA. However, one subject needed to be excluded from that analysis because he did not have any high confidence error trials (i.e., one of the cells in the 2 x 2 design was empty). To include that subject one could perform a simple t-test on the main effect of confidence and accuracy. This led to the same pattern of results for both confidence ($t(14) = 2.69$, $p = 0.02$) and accuracy ($t(14) = 0.55$, $p = 0.59$).

Since the present experiment is based on detection of coherent motion, we also examined the BOLD signal in motion sensitive area MT+. Although this area is sometimes considered to be part of the dorsal attention network, in the above analysis we delineated it as a distinct region in order to look at its effects separately. Unlike the effect in the dorsal attention network, pre-stimulus activity in MT+ was not associated with subjects' confidence ratings ($t(14) = 1.31$, $p = 0.21$). This lack of significant effect on confidence suggests the difference between high and low confidence trials in dorsal attention network was not driven by fluctuations of activity in MT+. Similarly, pre-stimulus activity in MT+ for error trials (mean activity = -0.03) and correct trials (mean activity = -0.01) was not significantly different ($t(14) = 0.39$, $p = 0.7$). Further, standard GLM analyses showed that for post-stimulus evoked activity, high activity in MT+ was associated with high level of confidence ratings ($t(14) = 2.98$, $p = 0.01$; Figure 23). Thus, in this study, the activity in MT+ played a bigger role in reflecting stimulus processing rather than pre-stimulus fluctuation of attentional states.

We further tested the relationship between confidence and pre-stimulus activity in two other networks. We found that pre-stimulus activity in neither the default mode network (Figure 24; Damoiseaux et al., 2006; Eichele et al., 2008; Fox et al., 2005; Sadaghiani et al., 2009) nor the alertness network (Figure 25; Sadaghiani et al., 2009, 2010) predicted confidence or accuracy on the task (all $p$'s > 0.1; Table 1).

The results on the dorsal attention network (DAN) corroborate our variance reduction model. One of the crucial assumptions of the VR model is that lack of attention increases the trial-by-trial variability of the perceptual signal. We provided an indirect test of this assumption by investigating whether pre-stimulus activity in DAN affected the variability in the evoked activity in MT+, a region that likely codes the perceptual signal. We computed the Fano factor of the distributions of evoked MT+ activity, a measure that has been used extensively to characterize neural variability (for example, see Carandini, 2004; Churchland et al., 2010; Gur et al., 1997). Confirming the VR model's prediction, high pre-stimulus activity in DAN led to a lower Fano factor in MT+ ($t(14) = 2.56$, $p = 0.02$; Figure 26). No such relationship was found for the default network ($t(14) = -0.71$, $p = 0.49$) or the alertness network ($t(14) = 0.4$, $p = 0.7$). We considered the alternative interpretation that the result for DAN was due to a ceiling effect: if the pre-stimulus BOLD activity is already high in DAN then pre-stimulus activity may be relatively high in MT+, and thus evoked activity in MT+ could potentially show ceiling effects. Thus we computed the skewness of the distributions of evoked activity in MT+. A ceiling effect would manifest itself as negative skewness. Nevertheless, we found that skewness was positive for trials with either low (skewness = 0.30) or high (skewness = 0.28) pre-stimulus DAN activity and there was no significant difference in skewness between the two distributions ($p = .92$).

Finally, in order to test whether the VR model in which attention decreases the variability of the perceptual signal can explain the observed data, we carried out simple computational simulations in which attention modulated linearly the standard deviation of the signal detection distributions (see Methods). These simulations were largely motivated by the fact that we did not observe a statistically significant effect of pre-stimulus BOLD signal in the dorsal attention network on motion discrimination accuracy, yet this was one of the predictions of the VR model. We performed the computational simulations four different

times and obtained good fits of the data all four times (Table 2 reports the fit of the first simulation that provided the best fit). Further, the parameter that controlled the attentional decrease of the variability of the perceptual signal was consistently positive across all four fits (see Methods) suggesting that attention indeed decreased perceptual noise (note that the parameter was not constrained and that it could have taken negative values). For the fitted values of the parameters we generated 10,000 replications of our experiment (by generating 15 sets of 100 trials for each replication). We compared the simulated pre-stimulus BOLD responses and found that simulated BOLD activity was significantly higher for low (compared to high) confidence trials on 7,188 trials (i.e., 72% of all simulations), while it was significantly higher for correct (compared to error) trials on 1,306 trials (i.e., 13% of all simulations). Thus our simulations suggest that there is a large difference in our statistical power to detect effects on confidence vs. effects on accuracy, which may explain why in our dataset only the effect on confidence was significant. Therefore, the lack of a significant effect of pre-stimulus BOLD signal in the dorsal attentional network on accuracy may well be due to the lack of statistical power (as estimated from these simulations), and thus does not necessarily contradict the variance reduction model.

**Discussion**

The study showed that low pre-stimulus dorsal attention network (DAN) activity was associated with high confidence in a motion discrimination task, as well as higher trial-by-trial variability in post-stimulus peak activity in MT+. The variance reduction (VR) model predicted this pattern of results though its last prediction – that low pre-stimulus DAN activity should lead to lower accuracy – was not confirmed.

One may worry that the finding of pre-stimulus BOLD in the dorsal attention network being negatively correlated with confidence could seem contradictory to previous single-

neuron recording research. For example, Kiani and Shadlen (2009) investigated the response of LIP neurons when monkeys were given the opportunity of choosing a safe option (thus indicating low confidence in their decision). The researchers found that the monkeys chose the safe option when the activity in LIP neurons was at an intermediate level and therefore the activity in the recorded neurons did not provide strong evidence for either decision option. Thus Kiani and Shadlen's results could be interpreted as predicting that the population response of LIP neurons would not distinguish between high and low confidence responses. Nevertheless, though Kiani and Shalden found neurons that commonly code for both accuracy and confidence, this finding does not necessarily imply that one cannot find dissociation elsewhere in the brain, or even within the same region. Also, Kiani and Shadlen's study was mainly concerned with post-stimulus neuronal activity, and did not investigate pre-stimulus activity, which is the focus of the current study. Finally, the relationship between measures of individual neurons and a population measure like fMRI can be complex, and higher activity for low compared to high confidence has indeed been reported in previous fMRI studies for parietal and frontal areas (Fleming, Huijgen, and Dolan, 2012). Thus, Kiani and Shadlen's work is not incompatible with the current results reporting higher pre-stimulus DAN activity for low confidence trials.

In a related study, Hesselmann, Kell, and Kleinschmidt (2008) investigated the influence of pre-stimulus activity in MT+ on the perception of coherent motion. They found that high pre-stimulus activity in right MT+ biases subjects to perceive coherent motion. The current study used motion that was always coherent and subjects simply needed to identify the direction of that motion. There were no significant differences in pre-stimulus activity in MT+ for error and correct trials. It is likely that pre-stimulus activity in MT influences the perception of coherence but does not necessarily lead to more accurate identification of the direction of motion.

Though the finding that attention decreased the variability of the sensory responses (Figure 26) is compatible with previous research (Mitchell et al., 2007; Bressler and Silver, 2010; Cohen and Maunsell, 2009), it needs to be interpreted with caution. The two categories of motion stimuli used in this study (contracting and expanding) are both expected to activate MT+. Thus, the average activity in MT+ does not directly reflect the evidence for one motion direction or the other. Also, BOLD activity is dominated by scanner and physiological noise (Fox and Raichle, 2007) that are not necessarily directly meaningful to the perceptual decision itself.

Another limitation of the study is that if the variance reduction model is correct, one may expect the increase in noise associated with lack of attention to be reflected by higher discrimination accuracy as well, something that was not observed in this study. On the other hand, the relative sizes of the impact of an increase in noise on accuracy and confidence may not be the same, and depend on factors such as how the confidence criteria are set. The study also had limited statistical power since due to the sparse event-related design employed each subject completed a limited number of trials (~100). Indeed the computational simulations suggest that even if the variance reduction model is correct, there may have nevertheless been limited statistical power to detect a significant positive association between pre-stimulus dorsal attention network activity and discrimination accuracy. The simulations were intended as a proof of concept and were not meant as a formal model of how attention influences pre-stimulus BOLD or the internal perceptual distributions. Nevertheless, these simulations demonstrate that the negative finding on the relationship between pre-stimulus BOLD activity on accuracy does not necessarily contradict the variance reduction model.

One needs to be cautious not to assume that pre-stimulus activity in the dorsal attention network can be absolutely equated with subject's attentional level. There are many other factors that contribute to activity in any large brain network (Fox and Raichle, 2007).

Nevertheless, the proposed interpretation of the current data is compatible with previous studies have shown that the brain modulates the attentional level by changing the state of cortical networks (Harris and Thiele, 2011) and that the dorsal attention network in particular reflects positively on one's attentional state (Eichele et al., 2008; Sapir et al., 2005).

Finally, even though one interpretation of the current results is that pre-stimulus activity in the dorsal attention network influences the perception of the stimulus, the data are also consistent with the interpretation that what is being influenced is a late, cognitive stage of the information processing. Nevertheless, we favor the former interpretation because in the studies reported in Chapters 1 and 2 attention not only led to lower visibility ratings but also to more conservative detection biases. Also, these effects were resistant to feedback and payoff manipulations, as if they were automatic and part of the perceptual processes themselves.

To sum up, the current results corroborate the psychophysical findings from Chapter 2 in that lack of attention may lead to liberal subjective biases such as higher subjective ratings. Such liberal subjective biases for unattended objects may partially explain why subjects find it surprising when they fail to perceive unattended objects (such as in inattentional and change blindness experiments) and can perhaps shed some light on why we think we see the whole visual scene in front of us, despite the fact that we seem to only be able to process a few objects effectively within our focus of attention (Kim and Blake, 2005).

# Chapter 5: Testing the Variance Reduction Model Using Transcranial Magnetic Stimulation

## Introduction

In Chapter 2 we showed that spatial attention boosts visual discrimination accuracy but lowers the corresponding visibility or confidence ratings. These data were nicely fit by the variance reduction (VR) model presented in Chapter 3. The critical assumption of the model was that lack of attention increased the trial-by-trial variability of the internal signal thus increasing the spread of the internal distributions. Some key assumptions of the model were tested using functional MRI in Chapter 4.

Interestingly, the VR model can easily generalize from attention to other processes that change the internal variability of the signal. According to the model, if the criteria for high confidence remain constant, an increase in signal variability should lead to a decrease in discrimination accuracy and an increase in confidence. To ensure that subjects do not consciously adjust their criteria to compensate for the change in signal variability, ideally the difference between the conditions of interest should be subtle and mainly focused on increasing variability in one of the conditions.

Though transcranial magnetic stimulation to the visual cortex is often used to completely 'knock out' conscious perception (e.g. Boyer et al., 2005; Breitmeyer et al., 2004; Kastner et al., 1998; Koivitso et al., 2010; Koivitso et al., 2011; Ro et al., 2004), some recent studies have shown that low intensity stimulation can effectively inject noise to the visual system (Ruzzoli et al., 2010; Schwarzkopf et al., 2011; though see Harris et al., 2008; Ruzzoli et al., 2011). Here we applied single-pulse TMS at intensity below the threshold for the conscious perception of phosphenes. The stimulus location was carefully adjusted to place the small target within the TMS-influenced region (as reflected by phosphene perception at

higher intensity). To anticipate, the results showed that single-pulse TMS to occipital cortex decreased discrimination accuracy but nevertheless increased confidence ratings. Confirming the variance reduction model, computational modeling showed that TMS seemed to influence mainly the variability but not the mean signal intensity of the visual percept.

## Methods

### *Subjects*

Eleven subjects participated in the experiment. Four of them did not perceive phosphenes under high intensity occipital stimulation with TMS. Because our task procedure critically depended on identifying the location of the effects of TMS within the visual field (see below), these subjects were excluded from the analysis. One subject performed the task at chance level and was also excluded from the analyses. This left us with six subjects (4 female, 19-40 years old). Subjects were required to have normal or corrected-to-normal vision. All subjects were screened with psychiatric, physical and neurological examinations, urine drug screens, and pregnancy tests for women of childbearing capacity. Potential subjects were excluded before the experiment if they had a history of current or past Axis I psychiatric disorder (including substance abuse/dependence) as determined by the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-NP), a history of neurological disease, or seizure risk factors. Subjects received detailed information about the potential side effects of TMS and signed an informed consent statement approved by the New York State Psychiatric Institute IRB.

### *Stimuli and task*

Subjects were seated in a dark room about 100 cm away from a computer monitor. They were required to fixate on a small white cross for the duration of the experiment (Figure

27). The task was to indicate whether a small bar (0.2º visual angle) was tilted 45 degrees to the left or to the right. The bar was presented for 33 ms (2 computer display refresh frames), and appeared just below the fixation cross. The exact location varied from subject to subject as we placed the bar at the location in the visual field where each subject experienced phosphenes from TMS. The small size of the bar was employed to ensure that the stimulus fell completely within the stimulated region of the visual cortex (Boyer et al., 2005; see description of TMS methods below). Stimuli were generated using Psychophysics Toolbox (Brainard, 1997) in MATLAB (MathWorks, Natick, MA) and were shown on a MacBook (13 inch monitor size, 1200 x 800 pixel resolution, 60 Hz refresh rate).

The fixation cross changed from white to black during the period when the bar was present. This change was the cue to subjects that the stimulus had appeared and that they needed to respond, guessing if necessary. Subjects responded by pressing one of four keyboard buttons that indicated both their decision about the tilt of the bar (left/right) as well their confidence in their decision (high/low). We did not use a more graded scale for confidence in order to keep the task as simple as possible and because a continuous confidence scale would have complicated our signal detection theoretic modeling (see below). Before the start of the experiment subjects were encouraged to use both levels of confidence as evenly as possible. This instruction was not repeated once the experiment started in order to avoid subjects deliberately adjusting their confidence ratings by employing cognitive strategies rather than perceptual processes. Trial duration during the TMS experiment was set at 5 seconds. We chose this relatively long interval to make sure that the TMS effect of the previous pulse had largely subsided by the time the next pulse was delivered, so that cumulative effects of TMS over the course of a block would be minimized.

In an initial training session, subjects practiced with the task over the course of 558 trials. This session occurred on a different day and in a separate room than the subsequent

TMS experiment. In the training session each trial began 500 ms after the response to the previous trial was given. Based on the training data, we determined for each subject three levels of luminance contrast for the tilted bar. A medium contrast level was chosen so that d' – a measurement of subjects' capacity to discriminate the orientation of the bar – would be between 1 and 2 (roughly corresponding to 70-85% correct responses). The lowest contrast was then fixed at 75% of the medium contrast, and the higher contrast was fixed at 125% of the medium contrast. We used three contrast levels because such variability in stimulus strength encourages subjects to use both sides of the confidence scale.

In the TMS session, the contrast of the bar was chosen pseudo-randomly on each trial so that the total number of presentations with each contrast was identical. For one subject (the first subject in our experiment) only the medium contrast level was used. Subjects were not explicitly informed about the presence of multiple contrast levels.

*TMS session*

Before the start of the TMS experiment, subjects were given 20 trials of practice with the task without TMS. This was done in order to familiarize the subjects with the TMS environment and to make sure that they remembered all the details of the task.

TMS was delivered with a figure-of-eight coil (9 cm diameter) powered by a Magstim 200 stimulator (Magstim Co., Whitland, South West Wales, UK). This coil provides weaker but more focused stimulation, compared to the round coil that was used in many previous studies that found complete suppression of visual stimuli (Amassian et al., 1989; Boyer et al., 2005; Corthout et al., 1999).

The precise location and intensity of the TMS was determined just prior to the beginning of the main experiment. We first used a hunting procedure to determine the optimal location for stimulation on the occipital cortex. We started by explaining to the subjects what

phosphenes are and how they could determine if their subjective experiences would be considered to be phosphenes. We then delivered a pulse of 70% of the maximal stimulator output to a location approximately 2 cm above the inion. The main axis of the coil was oriented parallel to the sagittal plane and the coil handle extended ventrally. If subjects failed to perceive a phosphene we stimulated around the initial site. If subjects still did not perceive any phosphenes, we increased the stimulation intensity by 10% and repeated the procedure. As mentioned above, four subjects did not perceive phosphenes even at 100% of stimulator output and were subsequently excluded from the experiment.

Once a subject perceived phosphenes, we moved the coil in 2 cm steps in lateral, medial, inferior, and superior directions until we found the place on the occipital cortex that resulted in the strongest and clearest phosphenes. The spot was then marked on the head and the rest of the experiment was done using that location.

We then proceeded to determine each individual's phosphene threshold in the following manner. Starting at 40% of the maximum stimulator output, we delivered three TMS pulses to the pre-determined location on the occipital cortex. If the subject did not perceive a phosphene, we increased the intensity by 5%. If the subject perceived a phosphene at least once, we continued to deliver TMS pulses until either 5 positive or 5 negative responses were given. If we received 5 negative responses first, we again increased the intensity by 5% and repeated the procedure. Once 5 positive responses were provided at a given intensity, that intensity level was chosen as the subject's "phosphene threshold." Afterward, all stimulation was delivered at 80% of the individual phosphene threshold. The resulting mean intensity of stimulation during the experiment was 52.5% (SD = 5.3%) of maximum stimulator output. We used jumps of 5% in order to minimize the time to find the phosphene threshold as in previous studies (Koivisto et al., 2010). However, it is important to note that this procedure may have slightly overestimated the phosphene threshold with up to

4%. Nevertheless, even with such overestimation our final stimulation intensity was lower than 90% of the real phosphene threshold. No subject reported seeing phosphenes or any other visual phenomena during the experiment; in fact, no subject noticed a difference in their perception of the visual stimuli. The intensity of TMS to the control site (vertex) was always the same as for the occipital cortex. No leg movement was elicited by vertex stimulation in any of the subjects.

The first subject completed 200 trials separated into 2 blocks, terminating the experiment after two blocks due to fatigue. The second subject completed 600 trials separated in 6 blocks, taking almost 3 hours to complete the experiment. The data from these two subjects did not differ in any systematic way from the data from the rest of the subjects. Subsequently we revised the total trial number and block size to prevent fatigue and excessive experiment duration. All subsequent subjects completed 468 trials separated into 6 blocks of 78 trials.

The TMS pulse was always delivered 100 ms after the onset of the bar stimulus. This interval was chosen based on a number of previous studies that found that TMS had maximum effect on visual perception between 80 and 120 ms after stimulus onset (Amassian et al., 1989; Boyer et al., 2005; Corthout et al., 1999; Kammer, 2007; Kastner et al., 1998; Luber et al., 2007; Maccabee et al., 1991; Miller et al., 1996). In half of the blocks subjects received a TMS pulse to the occipital cortex; in the other half of the blocks the subjects received a pulse at the same stimulus intensity to the vertex of the head. The vertex was an active control site, chosen because TMS applied there was not expected to interfere with the visual task, but controlled for the ancillary effects of TMS, including startle, acoustic artifact, and somatosensory sensation. The order of the occipital and vertex TMS blocks was counterbalanced across subjects.

After the experiment the subjects were debriefed about the purpose of the study and asked about the side effects of TMS. No subject reported more than mild discomfort from the TMS, and the discomfort did not continue after the end of the experiment. Visual analog scales were used to assess the subjects' emotional state before and after TMS and no significant differences were found.

*Data analysis*

The signal detection theoretic measure d' was calculated by first coding each trial as a hit, miss, false alarm, or a correct rejection. Trials in which subjects reported that the stimulus was right tilted were coded as hits if the bar was indeed right tilted and as false alarms if the bar was left tilted. Trials in which subjects reported that the stimulus was left tilted were coded as misses if the bar was right tilted and as correct rejections if the bar was left tilted. Hit rate (HR) was computed as hits / (hits + misses) and false alarm rate (FAR) was computed as false alarms / (false alarms + correct rejections). Finally, $d' = z(HR) - z(FAR)$.

*Computational modeling assumptions*

We fit the behavioral results using five different models. All models were based on signal detection theory (SDT). Two of them stipulated different channels for confidence ratings and stimulus discrimination. The other three were based on a single detection theoretic process.

In each model, we made standard assumptions: (1) the two stimuli used in the experiment gave rise to internal signals normally distributed along some decision axis; (2) perceptual decisions were made by comparing the signal on the decision axis to a criterion; (3) confidence judgments were made by comparing the signal on the decision axis to multiple criteria, corresponding to the confidence ratings available to subjects in this experiment; and

(4) criteria for perceptual decisions and confidence ratings were set in the same way for occipital and vertex TMS. The last assumption is justified for several reasons. First, none of the subjects reported that they had consciously perceived the task during occipital TMS to be any harder than during vertex stimulation. The relatively small difference in capacity we measured (see Results) also suggests that two conditions were not noticeably different to the subjects. Finally, previous research (Gorea and Sagi, 2000) has demonstrated that when lower and higher visibility stimuli are presented together in a block, subjects tend to use a single set of criteria for both. In the current experiment, the accuracy differences between the occipital and vertex TMS trials was closer than those found between conditions in Gorea and Sagi's experiment which makes it even less likely that subjects were able to shift their confidence criteria between them.

*Model specifications*

We compared between five possible models (Figure 30). The first three models postulated a single channel with two Gaussian distributions (corresponding to left- and right-tilted bars) with the difference in means $\mu$ and standard deviations of $\sigma$.

In Model 1 we allowed occipital TMS to only affect the distance ($\mu$) between the Gaussian distributions. The standard deviation $\sigma$ was set to an arbitrary constant value of 1; the numerical value of this parameter on its own is unimportant because its contribution is to be determined within the context of other parameters. The model had 5 free parameters: $\mu_{occipital}$, $\mu_{vertex}$, and the location of each of the three criteria levels used for discrimination and confidence judgments.

Model 2 tested if the primary effect of occipital TMS was inducing noise in the internal representations. It allowed occipital TMS to only affect $\sigma$ with no influence on $\mu$. The standard deviation for vertex TMS ($\sigma_{vertex}$) was again set to 1. The model had 5 free

parameters: $\mu$ (which had the same value for occipital and vertex TMS), $\sigma_{occipital}$ and the location of each of the three criteria levels used for discrimination and confidence judgments.

Model 3 allowed occipital TMS to affect both the signal and the noise of the internal representations. Thus, in this model occipital TMS changed both $\mu$ and $\sigma$. As in Model 2, the standard deviation for vertex TMS ($\sigma_{vertex}$) was set to 1. Overall the model included 6 free parameters: $\mu_{occipital}$, $\mu_{vertex}$, $\sigma_{occipital}$, and the location of each of the three criteria levels used for discrimination and confidence judgments.

The last two models included "conscious" and "unconscious" channels. In each of the two channels the left- and right-tilted bars gave rise to a Gaussian distribution. The distance between the centers of these Gaussian distributions were defined as $\mu_c$ and $\mu_u$ for the conscious and unconscious channels, respectively. Similarly, $\sigma_c$ and $\sigma_u$ were the standard deviations of the distributions for the conscious and unconscious channels, respectively.

Model 4 allowed occipital TMS to affect $\mu_c$ only. This is a model similar to previous accounts of the blindsight phenomenon (Boyer et al., 2005; Weiskrantz, 1986) that postulated that lesions to the primary visual cortex may affect a conscious cortical channel but leave intact an unconscious subcortical channel. Without loss of generality, we set $\sigma_c$ and $\sigma_u$ equal to 1. This left the model with six free parameters: $\mu_u$ (which had the same value for occipital and vertex TMS), $\mu_{c\text{-}occipital}$, $\mu_{c\text{-}vertex}$, and the location of each of the three criteria levels used for discrimination and confidence judgments.

However, one can argue that it is possible that the unconscious channel was also affected by occipital TMS. Therefore, in Model 5 we allowed occipital TMS to affect the unconscious channel as well. The model was equivalent to the first model in all other respects. It had 7 free parameters: the free parameter $\mu_u$ from Model 4 was now modeled as two parameters: $\mu_{u\text{-}occipital}$ and $\mu_{u\text{-}vertex}$, while the remaining 5 parameters were the same as in Model 4.

*Model fitting*

We fit the models to the data using a maximum likelihood estimation approach that has previously been used within a signal detection framework (Dorfman and Alf, 1969). Briefly, the likelihood of a set of SDT parameters given the observed data can be calculated using the multinomial model. Formally,

$$L(\theta|data) \propto \prod_{i,j} Prob_\theta(Resp_i|Stim_j)^{n_{data}(Resp_i|Stim_j)}$$

where each $Resp_i$ is a behavioral response a subject may produce on a given trial, and each $Stim_j$ is a type of stimulus that may be shown on a given trial. $Prob_\theta(Resp_i \mid Stim_j)$ denotes the probability with which the subject produces the response $Resp_i$ after being presented with $Stim_j$, according to the signal detection model specified with parameters θ. $n_{data}(Resp_i \mid Stim_j)$ is a count of how many times a subject actually produced $Resp_i$ after being shown $Stim_j$.

In the current study, the subjects had four possible responses from which to select (2 stimulus classification options * 2 levels of confidence), and there were two stimulus types (left- or right-tilted bars).

Note that the models were not fit to summary statistics of performance such as percent correct or average visibility. Rather, the models were fit to the full distribution of probabilities of each response type contingent on each stimulus type. From this full behavioral profile of stimulus-contingent response probabilities, we could derive various summary statistics.

We fit all models under consideration to the observed data by finding the maximum-likelihood parameter values θ. Maximum likelihood fits were found using a simulated annealing procedure (Kirkpatrick, Gelatt and Vecchi, 1983). Model fitting was conducted separately for each subject's data.

*Formal model comparison*

The maximum likelihood associated with each model characterizes how well that model captures patterns in the empirical data. However, comparing models directly in terms of likelihood can be misleading; greater model complexity can allow for tighter fits to the data but can also lead to undesirable levels of overfitting, i.e. the erroneous modeling of random variation in the data. The Akaike Information Criterion (AIC), motivated by considerations from information theory, provides a means for comparing models on the basis of their maximum likelihood fits to the data while correcting for model complexity (Burnham and Anderson, 2002). We used AICc, a variant of AIC which corrects for finite sample sizes:

$$AICc = -2*\log( L(\theta \mid data) ) + 2*K*(n / (n - K - 1) )$$

where K is the number of parameters in the model and n is the number of observations being fit. Lower values of AICc are desirable because they indicate a higher model likelihood and/or a lower model complexity (lower number of parameters).

We used the likelihood of each model, given the data, as a basis for model comparison:

$$L(model_i \mid data) \propto e^{-1/2 * (AICc\_i - AICc\_min)}$$

AICc_i is the AICc for model i and AICc_min is the lowest AICc exhibited by all models under consideration. These model likelihoods can be scaled to sum to 1, and the resulting "Akaike weights" reveal the relative weight of evidence for each model as evaluated by its fit to the data, correcting for model complexity.

We replicated the above analysis using the Baysian Information Criterion (BIC) in place of AICc, where

$$BIC = -2*\log( L(\theta \mid data) ) + K*\log(n)$$

The two different ways of performing model selection yielded similar results.

**Results**

We first analyzed our data using repeated-measures ANOVA in order to test for the effects of TMS site and contrast separately. Since our first subject only had one contrast level, that subject was excluded from the ANOVA. Using the remaining 5 subjects, we first analyzed the accuracy data (% correct) and found a significant main effect of TMS site ($F(1,4) = 14.01$, $p = .02$), a significant main effect of contrast ($F(2,3) = 10.31$, $p = .045$), and no interaction between TMS site and contrast ($F(2,3) = 0.62$, $p = .596$). We then analyzed the confidence data and again found a significant main effect of TMS site ($F(1,4) = 10.16$, $p = .033$), a significant main effect of contrast ($F(2,3) = 13.89$, $p = .03$), and no interaction between TMS site and contrast ($F(2,3) = 0.93$, $p = .485$). Thus occipital TMS has a significant influence on both accuracy and confidence. Since our main interest was in the effect of occipital vs. vertex TMS, and since contrast did not interact with TMS site, we pooled across contrast levels for each subject. This pooling allowed us to analyze the data from all 6 subjects simultaneously.

When we pooled across the three contrast levels and included all subjects, a paired sample t-test showed that compared to vertex TMS, occipital TMS decreased discrimination capacity d' ($t(5) = 4.49$, $p = .006$). The decrease was relatively small (0.16 on average, which corresponds to about a 2.5 percent decline in correct answers) but this was expected based on our procedure and choice of low intensity, localized stimulation (see Methods). Nevertheless, the result was consistent: it was present in each of our six subjects (Figure 28). Similarly, in

line with this detection theoretic analysis, the proportion of correct answers was lower for occipital TMS than vertex ($t(5) = 4.37$, $p = .007$). Also, occipital TMS did not bias subjects toward one stimulus type (right- or left-tilted bars; $t(5) = 0.63$, $p = .55$). On the other hand, confidence ratings increased after occipital TMS compared to vertex stimulation ($t(5) = 3.94$, $p = .01$). The effect was again consistent across our subjects with five of them showing an increase and one showing no difference in confidence between the two conditions (Figure 29).

In order to explain these somewhat counter-intuitive results, we fit 5 different models to the data (see Methods and Figure 30). The first three models postulated a single processing channel in which occipital TMS is posited to affect the signal only (Model 1), the noise only (Model 2), or the signal and the noise (Model 3), respectively. The other two models were dual-channel models that included 'conscious' and 'unconscious' channels that operate in parallel, with confidence rating being specifically dependent on the 'conscious' channel. In Model 4 occipital TMS affected the signal of the 'conscious' channel only, while in Model 5 occipital TMS affected the signal in both the 'conscious' and 'unconscious' channels.

Since the five models differed in complexity, we used Akaike Information Criterion (AIC) to compare the fit to observed data while punishing models for extra complexity (number of parameters). The results showed that the best fitting model was Model 2 in which occipital TMS affected only the variance of the internal representations (Figure 31). The model's average Akaike weight was two times higher than any of the other models. These results were confirmed by using a slightly different metric for punishing the complexity of the models, Bayesian Information Criterion (BIC; see Methods), which gave similar results with Model 2 emerging as the preferred model (model fits to the data are presented in Figure 32). Thus, our findings do not depend on the specific measure that we used to compare between the models.

We further investigated the parameter fits for each model. The best fit in Model 2 occurred when occipital TMS increased the standard deviation of the Gaussian distributions by about 12%. Further, all models included 3 free parameters corresponding to the discrimination criterion and the two confidence criteria. The average discrimination criterion was close to zero for all models which is in line with our finding that subjects were largely unbiased in discriminating between left- and right-tilted bars. The average values for the two confidence criteria were also virtually identical across the five models.

## Discussion

### *Summary of results*

The variance reduction (VR) model introduced in Chapter 3 predicts that higher variance in the perceptual signal, coupled with constant decision criteria, would lead to lower accuracy but higher confidence ratings. We tested this counter-intuitive prediction using low intensity TMS to the occipital cortex that affected the normal visual processing of a small tilted bar. As predicted by the VR model, the TMS decreased subjects' ability to identify the tilt of the bar while at the same time leading to higher confidence ratings. Further, formal model comparison confirmed that the main effect of TMS was in increasing the perceptual variability. These results provide further support for the VR model and demonstrate that it generalizes beyond the realm of attention.

### *Site of stimulation within the visual cortex*

It has sometimes been assumed that occipital TMS primarily targets the visual area V1 (Boyer et al., 2005; Laycock et al., 2007; Silvanto, Cowey, et al., 2005; Silvanto, Lavie, et al., 2005). However, a recent study (Thielscher et al., 2010) found that the exact site of stimulation is likely to be V2d or even V3. Nevertheless, a recent study demonstrated that

depending on each individual's occipital cortex folding, it could be possible to target V1 preferentially (Salminen-Vaparanta, Noreika, Revonsuo, Koivisto, and Vanni, 2012). Further, it is important to take into account that the proximity to the TMS coil is not the only factor that contributes to the effectiveness of TMS. It is well known that neurons are more excitable when the field is oriented along their axon and quite unexcitable to currents perpendicular to the axis (Rudin and Eisenman, 1954; Rushton, 1927). Consistent with this, some studies have found that the influence of TMS is greatest on regions whose cortical columns are longitudinally aligned to the direction of the current even if these regions are located sulcally while regions located in a gyrus – and thus physically closer to the coil – were relatively unaffected (Fox, Halko, Eldaief, and Pascual-Leone, in press). This is in agreement with a study that we report as Appendix B that suggested that V1 is more affected by offline TMS compared to V2 and V3. Thus, it is possible that the columnar organization of V1 makes it more excitable to occipital TMS even if the coil is positioned closer to other areas of the visual cortex. In any case, the current study does not critically depend on the exact regions in the visual hierarchy that were stimulated, since the net effect of increased noise could be accomplished at several levels in the hierarchy.

*Does TMS affect noise or mean signal intensity for visual perception?*

The current results, especially the computational modeling findings, suggest that single-pulse occipital TMS increases the noise of the internal representations in our discrimination task. This conclusion is in line with several studies that have argued that TMS acts by adding neural noise to the perceptual process rather by affecting signal strength (Ruzzoli et al., 2010; Schwarzkopf et al., 2011). Nevertheless, two other studies argued for the opposite conclusion (Harris et al., 2008; Ruzzoli et al., 2011). This suggests that the precise influence of TMS may depend on the specific stimulation and task parameters. Note

that the psychological noise increase observed in the current study can be the result of complex neuronal influences that are not necessarily described well as simple addition of random noise to neuronal firing. For example, psychological noise can be the result of TMS suppressing activity of certain subpopulation of neurons while increasing the activity in another (Siebner et al., 2009; Silvanto and Muggleton, 2008).

In a follow-up study (included here as Appendix B), we used theta-burst stimulation (TBS, Franca, Koch, Mochizuki, Huang, and Rothwell, 2006; Huang, Edwards, Rounis, Bhatia, and Rothwell, 2005) to suppress visual processing in an offline manner. We found that, unlike online single pulse TMS, theta-burst stimulation led to a signal strength decrease with no significant effect on signal variability.

These findings show that the effects of TMS are likely to be complex, and depend on specific stimulus and TMS parameters. A more systematic approach is needed to establish the precise effect of TMS for neuronal and psychological processes in different task contexts and under various stimulation procedures. Because of the complexity of this issue, the conclusion that TMS induced noise is only restricted to this study.

*Previous TMS studies that investigated confidence*

Several previous studies found that suprathreshold occipital TMS had detrimental effects on both confidence and visibility ratings. For example, Boyer et al. (2005) found that TMS to the occipital cortex led to a substantial amount of trials in which subjects were unaware of the orientation (61%) or the color (70%) of the stimulus. Nevertheless, in these trials subjects performed better than chance in guessing the orientation or color, respectively. Even though this study suggests that objective performance may go beyond what subjects experienced subjectively after TMS, the different methodology makes it hard to directly compare Boyer et al.'s results with the results reported here. In particular, the purpose of their

study was to abolish the conscious percept, and they used stronger stimulation intensities and a circular coil, which typically leads to more intense but less focal effects. The present study, on the other hand, aimed to deliver localized and low intensity stimulation, by using a figure-of-eight coil and intensity below the threshold for phosphene perception. And yet, interestingly, when only the low confidence trials in the current study are considered, subjects also performed better than chance (66% correct responses) which parallels the results of Boyer et al.

In another study, Koivitso et al. (2011) found that suprathreshold TMS led to a decrease in both accuracy and subjective awareness, the latter of which is usually considered to be akin to confidence ratings (Szczepanowski and Pessoa, 2007). However, in that study the researchers achieved a much higher level of suppression – accuracy decreased by about 20% compared to the no-TMS condition. The single-channel SDT explanation is likely to be restricted to small changes in variance that is not subjectively noticeable (such that subjects do not consciously adjust their criteria for giving high confidence responses). Thus, it is perhaps not surprising that the variance reduction model would not apply to the results of Koivitso et al. (2011) where one may expect that the large decrease in accuracy led subjects to consciously adjust their confidence ratings. In comparison, in the present experiment, accuracy only decreased by 2.5%.

Finally, Koivitso et al. (2010) used subthreshold TMS but still found decreases in both accuracy and visibility ratings with occipital TMS. However, in that study the authors used motion stimuli for which processing depends on a different visual area (MT/V5). A further difference with the current study was the fact that the motion stimulus was several times larger than the small bar that we used which was designed to fall completely within the region where subjects perceived phosphene upon TMS. Finally, it is possible that despite their strong

association (Szczepanowski and Pessoa, 2007), visibility and confidence ratings may dissociate in the context of occipital TMS.

*Effect of subthreshold TMS on accuracy*

As noted above, Koivitso et al. (2010) found that subthreshold TMS to the occipital cortex can lead to a decrease in accuracy. Nevertheless, two other studies (Abrahamyan et al., 2011; Schwarzkopf et al., 2011) found an *increase* in accuracy after subthreshold TMS. Unlike the present study, Schwarzkopf et al. (2011) used motion stimuli and targeted MT rather than the primary visual cortex, and delivered a triple-pulse TMS (pulse gap of 50 ms). Abrahamyan et al. (2011) used much larger stimuli than in the current study (6.5º compared to 0.2º visual angle in our study; a difference of over 30 times) and employed a two-interval forced choice detection task which requires very different neural computation compared to the presently used single-stimulus discrimination task (Macmillan and Creelman, 2005). Overall, the differences between the current study and the previous studies (Abrahamyan et al., 2011; Koivitso et al., 2010; Schwarzkopf et al., 2011) prevent any conclusions on the general effects of subthreshold TMS on accuracy as these effects likely depend on factors such as size and type of stimulus used, stimulation site, stimulation procedure, and precise stimulation intensity. Therefore, the present data should not be considered as evidence that subthreshold TMS has a general detrimental effect on performance outside of the paradigm used here. More systematic research is needed to pinpoint the exact effects of subthreshold TMS and how they depend on the factors mentioned above.

*Conclusion*

We tested whether the VR model would generalize beyond the realm of attention. We used TMS to directly induce noise in the perceptual signal. As predicted by our VR model,

inducing noise with occipital TMS decreased accuracy but increased confidence ratings. Therefore, these results demonstrate the flexibility of the VR model and its ability to generalize to situations that do not involve attentional manipulations.

# Chapter 6: Attention Leads to Improved Metacognitive Sensitivity

## Introduction

The present work has thus far demonstrated that lack of attention leads to inflated subjective perception, as revealed by higher subjective ratings in various discrimination tasks (see Chapters 2, 4, and 5). However, to this point we have only been concerned with the influence of attention on the overall propensity of the subject to use high or low subjective ratings. What remains unclear is whether attention affects the informational content of these subjective ratings. Specifically, does attending to a stimulus make the subsequent subjective rating more or less predictive of subject's accuracy?

The ability of subjective ratings to predict one's accuracy is one of the important reasons why subjective ratings are important in everyday life (Lau, 2007; Metcalfe and Shimamura, 1994). For example, we expect that a witness that has a high degree of confidence in her report is more likely to be correct than another witness that is less confident. Similarly, the degree of confidence a physician places in her diagnosis would determine whether a risky operation needs to be performed.

The capacity with which subjects' subjective ratings predict their accuracy is sometimes refered to as "metacognition" (Cleeremans, Timmermans, and Pasquali, 2007). We investigated the attentional influence on metacognitive sensitivity using a similar design as Experiment 2.1. To anticipate, we found that attention improves metacognition, despite its negative influence on the overall propensity to use high subjective ratings. We introduce the "variance and criterion jitter reduction" (VCJR) model – a simple extension of the VR model – that accounts for the present findings by introducing trial-by-trial variability in the decision criteria that is larger for unattended stimuli.

## Methods

### *Participants*

Twenty-two subjects in total (14 women; age range = 19-32 years) participated in the experiment. All subjects were naive regarding the purposes of the experiments, had normal or corrected-to-normal vision, and signed an informed-consent statement approved by the local ethics committee.

### *Materials and procedure*

The experiment was similar to Experiment 2.1. The only difference was that subjects were given four different visibility ratings to choose from. In comparison, Experiment 2.1 only had two visibility ratings which makes the estimation of metacognitive score less reliable (Macmillan and Creelman, 2005; Maniscalco and Lau, 2012). We also increased the number of participants in order to have larger statistical power to detect any effects on metacognitive sensitivity.

As in Experiment 2.1, subjects completed 960 experimental trials separated into 4 runs of 6 blocks. Feedback was given at the end of each block consisting of 40 trials. Five subjects needed to be excluded: four due to inability to perform better than chance and one due to using almost exclusively the lowest level of visibility which made it impossible to compute reliably her metacognitive score using any of the three measures introduced below.

### *Measures of metacognition*

Discriminating between one's own correct and incorrect responses with subjective ratings has been called a ''Type 2 task'' (Clarke, Birdsall, and Tanner, 1959; Galvin, Podd, Drga, and Whitmore, 2003), as opposed to the ''Type 1 task'' of discriminating between

stimulus alternatives. Type 2 task performance is also referred to as metacognitive sensitivity (Maniscalco and Lau, 2012).

Metacognition has traditionally been measured by assessing the correlation across trials between the subjective rating and whether or not a correct answer was given in each trial (e.g., the measure "phi" in Kornell, Son, and Terrace, 2007, and the measure "gamma" in Nelson, 1984). However, while this approach is appealing because of its simplicity and intuitiveness, it risks introducing biases in certain situations. For example, the magnitude of the correlation can be influenced by the overall propensity of a subject to use high or low subjective ratings, as well as by their performance at the Type 1 task (Maniscalco and Lau, 2012).

An alternative approach to estimating metacognitive sensitivity has been used in several other studies (Fleming et al., 2010, 2012; Kolb and Braun, 1995; Wilimzig, Tsuchiya, Fahle, Einhäuser, and Koch, 2008). The idea behind it is to construct a Type 2 receiver operating characteristics (ROC) function. The Type 1 ROC is simply a plot of multiple pairs of false alarm rate (FAR) and hit rate (HR) constructed by asking an observer to provide a stimulus classification and a subjective rating on every trial (Macmillan and Creelman, 2005). To construct the Type 2 ROC, Type 2 hits and misses are defined as correct trials that receive high and low subjective ratings, respectively, while Type 2 false alarms and correct rejections are defined to be incorrect trials that receive high or low subjective ratings, respectively. Using these definitions, it is possible to construct a Type 2 ROC plot of the Type 2 FAR and HR. When multiple subjective ratings are collected, one can just connect the different points on the Type 2 ROC plot to construct a Type 2 ROC curve. The area under this curve can be used as a measure of metacognitive sensitivity (named $A_{roc}$ in Fleming et al., 2010, 2012, and $A_g$ in Pollack and Hsieh, 1969).

Although the above approach does not suffer from a bias due to subjects' overall propensity of using high or low subjective ratings, the resulting metacognitive estimate is contaminated by Type 1 performance. For example, a person who has higher d' in a task will naturally have higher area under the Type 2 ROC curve and will therefore appear to have higher metacognitive sensitivity (Clifford, Arabzadeh, and Harris 2008; Galvin et al., 2003; Maniscalco and Lau, 2012). This is not a concern in the current study since we ensured that d' for cued and uncued stimuli is equated, and therefore the measure $A_{roc}$ is appropriate here.

Finally, Maniscalco and Lau (2012) recently introduced a new measure of metacognitive sensitivity called meta-d'. The measure is based on signal detection theoretic (SDT) assumptions. Meta-d' characterizes observed Type 2 sensitivity as the value of d' that a metacognitively optimal observer would have required to produce the empirically observed Type 2 data. Thus, a meta-d' of 1 would signify that the Type 2 data observed would be produced by an ideal observer with d' of 1. On the other hand, the actual d' observed in the experiment could be lower, higher, or equal to 1. For cases in which one wants to compare the metacognitive score across conditions that produced different levels of d', one can use the measure M = meta-d' – d' (Maniscalco and Lau, 2012). Here we used directly meta-d' since d' values were matched between the cued and uncued conditions.

*VCJR model fitting*

We attempted to construct a model to explain the finding that attention improved metacognitive sensitivity. Our VR model could not account for this finding because the model postulates that subjects fix their criteria throughout the experiment and this results in optimal metacognitive performance such that meta-d' = d' (Maniscalco and Lau, 2012). Since in our dataset meta-d' was substantially lower than d' for most subjects, our VR model was unable to explain these data.

Therefore, we introduced an additional feature to the VR model: the decision criteria were allowed to jitter from trial to trial. Such criterion jitter results in suboptimal metacognitive performance so that meta-d' < d'. We called the new model "variance and criterion jitter reduction" (VCJR) model.

In order to optimize the computational demands for the model fitting, we did not fit the values of meta-d'. Instead, we used the same model fitting procedure as for Experiment 2.1 (see Chapter 3) to get estimates of the parameters controlling the standard deviation and the mean separation of the signal detection distributions for cued and uncued trials. We then introduced criterion jitter of different amounts for the cued and uncued trials separately, and checked what level of criterion jitter fits best the $A_{roc}$ data for the cued and uncued stimuli. We expected that there would be higher criterion jitter for the uncued stimuli but in order to avoid biasing the results, we allowed criterion jitter to vary freely for both the cued and uncued conditions.

Thus, our implementation of the VCJR model included 12 free parameters. The first 10 parameters paralleled parameters used in Chapter 3 to model Experiment 2.1: $\sigma_{cued}$, $\mu_{uncued}$, $\mu_{cued}$, and the criteria levels used for discrimination and visibility judgments (7 parameters). The last two parameters – $j_{cued}$ and $j_{uncued}$ – quantified the amount of criterion jitter present in the cued and uncued trials, respectively.

*Estimating criterion jitter in the observed data*

We used two methods to indirectly assess the amount of criterion jitter in the cued and uncued conditions. Firstly, we investigated if the decision criteria drifted over the course of the experiment. In separate analyses, we split the experiment into 2, 3, 4, or 5 equal epochs and estimated the criteria used for the cued and uncued trials in each epoch. This standard analysis assumes that the standard deviations of the distributions did not change from one

epoch to another. We then compared the congruency of the four sets of criteria values for the cued and uncued conditions. To do that, we computed the absolute value of the difference in the decision criteria for each pair of epochs. The total criterion drift was computed as the sum of criterion difference in all pairs of epochs.

Secondly, we investigated the amount of trial-to-trial dependencies by computing the criteria for both cued and uncued trials after subjects gave a particular subjective rating on the previous trial (1 to 4). As above, we then computed the absolute value of the difference in the decision criteria for each possible subjective rating on the previous trial. The total amount of trial-to-trial dependencies was computed as the sum of criterion difference in all pairs of possible subjective ratings on the previous trial.

*Statistics*

As in Experiment 2.1, most tests that we performed were paired-sample t-tests. In order to compute the signal detection measure d' we again first coded each trial into a hit, miss, correct rejection, or a false alarm.

**Results**

We investigated whether attention affects subjects' metacognitive sensitivity. We first checked whether we were able to replicate the results of Experiment 2.1. Indeed, we found that d' was successfully equated between the cued and uncued stimuli ($t(16) = 0.97$, $p = .35$; Figure 33a), while at the same time the visibility ratings were higher for the uncued (mean = 2.63) compared to the cued (mean = 2.32) stimuli ($t(16) = -2.62$, $p = .018$; Figure 33b). Therefore, the basic findings of Experiment 2.1 were fully replicated.

We also checked that the cued stimuli indeed benefited from the increased attention. The mean contrast value used for the cued stimuli (9.8%) was lower than that for the uncued

stimuli (15.5%, $t(16) = 2.11$, $p = .05$). Further, subjects responded to the cued stimuli faster than to the uncued stimuli ($t(16) = 4.81$, $p = .0002$; Figure 33c). Together, these findings demonstrate that cuing indeed led to behavioral improvements.

Therefore, we next turned to the critical issue in this study: whether attention led to a change in metacognitive sensitivity. To be conservative, we used three different measures of metacognitive sensitivity: meta-d' (Maniscalco and Lau, 2012; Rounis, Maniscalco, Rothwell, Passingham, and Lau, 2010), $A_{roc}$ (the area under the Type 2 ROC curve, see Fleming et al., 2010, 2012; Kolb and Braun, 1995; Wilimzig et al., 2008), and phi (the correlation between accuracy and subjective ratings, see Kornell et al., 2007).

The results showed that attention increased meta-d' ($t(16) = 2.7$, $p = .016$; Figure 34a), $A_{roc}$ ($t(16) = 2.53$, $p = .022$; Figure 34b), and phi ($t(16) = 3.01$, $p = .008$; Figure 34c). The congruity of these findings across three different measures of metacognition provides strong evidence that attention indeed strengthened metacognitive sensitivity.

How did attention improve metacognition? Our VR model could not account for the data because it postulates that subjects have fixed criteria over the course of the experiment. Such fixed criteria result in optimal metacognitive performance so that meta-d' = d', and consequently, according to the VR model, meta-d' should be equal for the cued and uncued conditions in this experiment (since d' was equal for the two conditions).

Here we introduce an extension of the VR model that postulates that the criteria for subjective ratings jitter from trial to trial (Mueller and Weidemann, 2008). This criterion jitter leads to suboptimal metacognitive performance. To account for our findings, the model postulates that criterion jitter is larger for the uncued compared to the cued stimuli. Therefore, we call the new model the "variance and criterion jitter reduction" (VCJR) model of attention to highlight the fact that attention decreases both the variability of the signal and the criterion jitter.

We implemented the VCJR model computationally by fitting it to the data (see Methods). The model was able to provide a good fit to the data (see Figure 35 where we have plotted the fitted d', visibility ratings, and $A_{roc}$ values). Further, we examined the parameter fits for the amount of criterion jitter. It turned out that, as we expected, criterion jitter was higher for the uncued stimuli compared to the cued stimuli (that is, $j_{unued} > j_{cued}$, see Methods; $t(16) = 2.68$, $p = 0.016$).

The critical feature of the VCJR model is that criterion jitter is larger for the uncued compared to the cued stimuli. There is currently no method to directly assess the total amount of criterion jitter in an experiment (Mueller and Weidemann, 2008). However, there are two components of criterion jitter that can indeed be quantified.

Firstly, the criteria may drift over the course of the experiment. Even though such a process does not necessarily imply that the criterion is oscillating from trial to trial, the effect of such a process over the course of the experiment would mimic the effect of trial-to-trial jitter. We therefore explored the extent to which the criteria for cued and uncued trials drifted over the course of the experiment. We split the experiment into four quartiles and estimated the criteria used for the cued and uncued trials in each of them. We then compared the congruency of the four sets of criteria values for the cued and uncued conditions (see Methods). We found that the uncued trials exhibited less congruency ($t(16) = 3.8$, $p = .002$). We repeated this analysis by splitting the experiment into 2, 3, and 5 epochs and obtained similar results for all of them (all $p$'s $< .05$). However, since the criteria for the cued condition were estimated based on a higher number of trials, it is possible that our results were due to the higher stability of the estimation for cued trials. To check that, we took a subset of the cued trials for each epoch so that we had equal number of cued and uncued trials for the estimation of the criteria. Since a single subset of trials may produce biased results, we repeated the above procedure one thousand times. We performed these analyses by splitting

the experiment again into 2, 3, 4, or 5 epochs and still found that across the one thousand simulations for each splitting, cued trials exhibited less criterion drift (all $p$'s < .05).

Secondly, it is known that, in most psychophysical tasks, subjects exhibit trial-to-trial dependencies such that the response on the current trial depends on the response given on the previous trial (DeCarlo and Cross, 1990; Jesteadt, Luce, and Green, 1977; Mori, 1998; Mueller and Weidemann, 2008; Ward, 1979). Over the course of an experiment, such dependencies also mimic the effect of increased criterion jitter. As above, we computed the criteria for both cued and uncued trials after subjects gave a particular subjective rating (1 to 4) on the previous trial. We found that the congruency of the criteria was larger for cued compared to uncued trials ($t(16) = 4.33$, $p = .0005$) demonstrating that uncued trials exhibited larger trial-to-trial dependencies. Since we had higher number of trials for the cued stimuli, we again ran one thousand simulations in which we only considered a subset of the cued trials so that there was equal number of cued and uncued trials. Again, despite using the same number of trials for the cued and uncued conditions, we still found larger trial-to-trial dependencies for the uncued stimuli ($p < .05$).

**Discussion**

The present study demonstrated that attention improves metacognitive sensitivity. The finding was robust across several different measures of metacognition. These results were explained by our "variance and criterion jitter reduction" (VCJR) model of attention which postulates that attention decreases both the variability of the perceptual signal and the jittering of the decision criteria.

To the best of our knowledge, this is the first demonstration that endogenous attention in a visual task improves metacognition. Note that in most psychophysical tasks, attention improves performance (i.e., capacity d'). Therefore, as noted in the Methods, such studies are

not well suited to address the question of metacognition since most measures of metacognitive sensitivity are biased by the performance on the main task (one notable exception is the new method of estimating metacognitive sensitivity introduced in Maniscalco and Lau, 2012). To avoid this issue, here we used low contrast cued stimuli and high contrast uncued stimuli to ensure that the performance at the main task (discriminating between right and left tilt stimuli) was equated.

Previous empirical work has linked metacognitive capacity with the prefrontal cortex (Fleming et al., 2010, 2012), while theoretical work has suggested links between attention and metacognition due to their reliance on prefrontal areas (Fernandez-Duque, Baird, and Posner, 2000). Our results confirm the existence of such a link in visual perception and suggest a mechanism for it. According to our VCJR model, the critical component that leads to the observed attentional improvement of metacognition is that attention is able to keep the decision criteria more stable over time. Partial support for this assumption was provided by our analyses of criterion change over the course of the experiment, as well as the presence of trial-to-trial dependencies. Therefore our data suggest that the critical involvement of the prefrontal regions may be in keeping consistent decision criteria over time. Such proposition is in line with previous theoretical work that has suggested a central role for frontal regions in criterion setting (Lau, 2007).

# General Discussion

## Summary of findings

The relationship between attention and the other components of perceptual decision making has come to the forefront of vision science research (Carrasco, 2011). In particular, a number of empirical and theoretical works have tried to relate attention and consciousness, as well as attention and appearance. However, what has been missing from this literature is an investigation of the attentional influence on the subjective aspects of perception for detecting and discriminating attended and unattended stimuli.

This dissertation attempts to fill in this gap. In particular, we set out to investigate how attention affects the detection criterion (Question 1), how attention influences subjective ratings (Question 2), what mechanisms could account for these attentional influences (Question 3), and whether attention affects metacognitive sensitivity (Question 4).

Results from all studies showed that attention leads to conservative subjective biases. In particular, we found that subjects adopted a conservative detection criterion for attended stimuli: they reported cued stimuli as appearing less often that uncued ones (Chapter 1). This effect was extended by the finding that attention also led to low subjective ratings in discrimination experiments (Chapter 2). In both sets of experiments the results were strongest when capacity d' was equated between different levels of attention, or when stimuli had low contrast. These results were explained by our variance reduction (VR) model of attention, which postulates that attention decreases the variability of the perceptual signal while keeping the decision criteria constant (Chapter 3). This model was subsequently tested with a functional magnetic resonance imaging (fMRI) study (Chapter 4). The study showed that a high attentional state, as indicated by high activity in the dorsal attention network in the brain, led to lower confidence ratings and lower variability in the evoked signal in motion sensitive

area MT+, thus supporting two key predictions of the model. The VR model was further tested by extending it beyond the realm of attention (Chapter 5). We used transcranial magnetic stimulation (TMS) to directly increase the variability of the perceptual signal. The effects mirrored the effect of lack of attention: TMS led to decreased performance but increased subjective ratings. Finally, we investigated the influence of attention on metacognitive sensitivity and found that despite the fact that attention decreased the overall magnitude of the subjective ratings, it made them more predictive of subjects' accuracy (i.e., it increased metacognitive sensitivity; Chapter 6). To account for these findings, a simple extension to the VR model – the "variance and criterion jitter reduction" (VCJR) model – was introduced wherein the decision criteria were allowed to jitter more for the uncued compared to the cued stimuli. We discuss the implications of these findings below.

**Relation to previous research**

To the best of our knowledge, these results are the first to demonstrate that under certain conditions, attention can make subjects more conservative or decrease the visibility of stimuli. This is especially surprising since, in most psychophysical tasks, there is typically a positive association between confidence and accuracy (Busey et al., 2000; Fleet et al., 1987).

Several previous studies reported findings that are congruent with our results here. Downing (1988) found that subjects had optimal detection criterion for uncued stimuli but were too conservative in detecting cued stimuli. However, in that experiment the probability of occurrence of the uncued stimuli was 50% while the probability of occurrence of the cued stimuli was 80%. It has been shown that subjects do not shift their criterion far enough when the probability of the stimuli is not 50% (Macmillan and Creelman, 2005). Thus, Downing's results could be attributed to the change in the prior probability of the stimuli and indeed this

is the interpretation that the author gave to his results. Here we employed unbiased priors and consequently we attribute our results to perceptual rather than cognitive processes.

Another study (Wilimzig et al., 2008) reported results congruent to our findings from Chapter 2. The researchers measured the difference in detection capacity d' and confidence for strongly and weakly attended stimuli. They found that attention boosted detection capacity d' but either did not affect confidence ratings at all or increased them by a small non-significant amount and these results were not modulated by the difficulty of the task. On the other hand, here we report lower visibility ratings for strongly attended stimuli but only when contrast is low. Further, unlike Wilimzig et al., we find a similar pattern of results when we equate capacity d' values between attentional conditions.

In a different study Baldassi, Megna, and Burr (2006) showed that when attentional focus was distributed to a larger set of stimuli, subjects made errors with higher confidence. Despite the superficial similarity with our results, the design of that study was significantly different and its conclusions do not generalize to situations where the targets are well defined in advance.

On the other hand, another study reported results seemingly opposite to what we find in Chapter 1. In their Experiment 3, Prinzmetal and colleagues (2008) reported that exogenous attention oriented by the presentation of a brief cue made subjects more liberal in a detection task. Surprisingly, in that study attention also lowered the detection capacity d'. Thus, it is possible, as the authors also suggest, that the cue interacted in some way with the stimulus thus decreasing detection capacity d' and making subjects more liberal.

There have also been several studies that reported dissociations between accuracy and confidence outside of the context of attention. For example, Lau and Passingham (2006) used presented a brief prime that was afterward masked by the stimulus that needed to be discriminated. The researchers found that at two different stimulus onset asynchronies (SOA)

between the prime and the stimulus, subjects had the same level of accuracy but gave significantly different confidence ratings. Dissociation between accuracy and confidence has also been demonstrated in patients with lesions of the primary visual cortex that resulted in a condition known as “blindsight” (Weiskrantz, 1986).

The study that we reported in Chapter 4 also suggests an alternative interpretation of a previous study by Sadaghiani et al. (2009) who reported that low pre-stimulus dorsal attention network (DAN) activity led to more hits in an auditory detection task. Sadaghiani et al.’s findings may suggest that low level of attention improves detection capacity, at least for auditory stimuli. However, the increase in detection rate can be attributed to either an enhancement of capacity, or a shift of detection criterion towards the liberal direction (responding “yes” more often). Compatible with this interpretation, in Chapter 2 we reported empirical data that demonstrated that there is a close relationship between high subjective perceptual ratings and liberal detection criteria. Thus the result that high pre-stimulus activity in DAN leads to low confidence ratings suggests that Sadaghiani et al.’s findings may be partially explained by a liberal detection bias caused by lack of attention.

Alternatively, Sadaghiani and colleagues (2009) suggested that the negative relationship between pre-stimulus DAN activity and detection performance may be due to the fact that sounds are not spatial (but see Tark and Curtis, 2009). In contrast with the current results, in their study DAN did not show considerable evoked activity. Given the differences in task (discrimination vs. detection) and stimuli used (motion patches vs. auditory tones) between the current study and that of Sadaghiani et al. (2009), it is possible that the influence of DAN depends on the specific context of the experiment as well. It is important to note that our detection theoretic account of Sadaghiani et al.’s findings, and their own interpretation, are not necessarily mutually exclusive.

**Mechanisms for confidence ratings**

One of the most startling findings reported here is the consistent dissociation between accuracy and subjective ratings across a variety of studies. Such dissociations have often been explained by dual-channel models (Del Cul et al., 2009; Jacoby, 1991; Jolij and Lamme, 2005; Morewedge and Kahneman, 2010). In dual-channel models, typically one channel supports conscious processing (this is often assumed to be a cortical channel that goes through the primary visual cortex) while the other is largely unconscious (a sub-cortical channel that bypasses the visual cortex). Thus, within the context of experiments involving subjective ratings, only the 'conscious' channel contributes to confidence ratings while both channels may contribute to the perceptual decision. It is easy to see how dual-channel models can account for dissociations between confidence and accuracy. For example, the phenomenon blindsight (Weiskrantz, 1986) could be due to a disruption in the 'conscious' channel (thus resulting in a lack of conscious visual experience) while the remaining visual processing would be based on the 'unconscious' channel that is left largely intact.

The VR model presents an alternative explanation for dissociations between accuracy and subjective ratings that is based on a single channel account. In fact, in Chapter 5 we used a formal model comparison analysis to show that our implementation of the VR model fit the data better and more parsimoniously compared to dual channel models. We do not challenge that dual-channel models are plausible in some contexts. However, in the present studies, the VR model is able to parsimoniously account for the data without introducing additional processing channels.

If the VR model is a faithful description of the computations performed in the brain, then one possible conclusion is that all relevant activity takes place within the visual cortex. Nevertheless, the fact that the VR model postulates that both accuracy and subjective ratings are based on the same signal does not necessarily imply a lack of involvement of higher areas.

In fact, a critical aspect of our model is that high confidence is generated when the signal strength is large enough to cross the relevant criterion, and that this process depends on how the criterion is set. Previously, it has been argued that such criterion-setting mechanism may depend on the prefrontal cortex (Lau and Rosenthal, 2011). This idea has received empirical support from studies that have demonstrated that metacognitive sensitivity in visual tasks depends on the frontal cortex (Fleming et al., 2010; 2012). Therefore, we believe that despite the fact that all relevant computations are likely performed on the same signal, these computations are distributed over several brain regions and are not localized to the visual cortex.

It is also important to note that the VR model is not intended as a general mechanism about how attention, accuracy, and subjective ratings are related since this relationship is likely complex and could depend on the exact setting of the experiment (Kiani and Shadlen, 2009; Reynolds and Heeger, 2009; Wilimzig et al., 2008). It is highly likely that generating confidence ratings in different contexts, such as medical decisions or witness testimonies, depend on additional mechanisms that are not captured by the VR model.

**The assumptions of the VR model**

*Constant decision criteria*

One of the central features of the VR model is that the criteria for detection, discrimination, and subjective ratings are the same for high and low attentional conditions (or, in the case of the TMS study, occipital and vertex TMS). This is a strong assumption, especially since it postulates a glaring suboptimality in the human visual system.

There are several reasons why we think that this assumption is reasonable. Most importantly, much previous research (Gorea, Caetta, and Sagi, 2005; Gorea and Sagi, 2000, 2001; Zak, Katkov, Gorea, and Sagi, in press) has demonstrated that when lower and higher

visibility stimuli are presented together in a block, subjects tend to use a single set of criteria for both. In these experiments, subjects were informed in advance as to the contrast levels of each stimulus, and thus they could have easily used two separate sets of decision criteria for the two types of stimuli. The fact that all of these studies found that subjects were unable to maintain two separate sets of decision criteria argues strongly for a general limitation in perceptual decision making when more than one stimulus is presented within a block.

Furthermore, even in Experiment 1.6, where contrasts were blocked and subjects could have shifted their decision criterion over blocks, they did not appear to have done so. Indeed, the false alarm rate was constant for both the cued and uncued stimuli (see Figure 8) suggesting that subjects fixed their detection criterion with respect to the noise distribution and did not significantly shift it between blocks.

The constant criterion assumption is potentially a concern in the TMS experiment (Chapter 5) because occipital and vertex TMS were delivered in different blocks of trials, and subjects had ample chance to shift their confidence criteria. Nevertheless, there are several reasons why we believe that this is unlikely. Firstly, all of our subjects reported that they did not notice a difference in the appearance of the stimuli in the occipital and vertex stimulation conditions. Many of them even commented that they were certain that occipital TMS did not have any influence on them. Thus, it is likely that they did not make a conscious effort to make any adjustments to their confidence judgments between the occipital and vertex stimulation conditions in the experiment. Secondly, several aspects of the data were also incompatible with the interpretation that subjects deliberately adjusted their confidence criteria. As can be seen in Figure 29, three of the six subjects used the confidence levels in a very biased manner (overusing either the high or low confidence responses) which suggests that they did not continuously re-adjust their confidence criteria to achieve somewhat even usage of high and low confidence ratings. Additionally, if such deliberate confidence

adjustment actually occurred, one would expect that confidence would decrease for vertex TMS and increase for occipital TMS over the course of the experiment. The reason for this is that adjustments to the confidence criteria would become easier later in the experiment when subjects have seen trials with both occipital and vertex TMS. However, our results showed that the difference in confidence was highest between the first vertex and first occipital TMS blocks (8.6% more high confidence trials in the occipital TMS block) compared to the second (2.1% difference) and third (3.6% difference) blocks. This pattern of results is the opposite of what one would expect if the difference in confidence was due to subjects deliberately adjusting their confidence criteria. Finally, if subjects adjusted their confidence ratings in the TMS experiment, one would predict that confidence should be the same across the occipital and vertex stimulation conditions. It is hard to see why in the case of occipital stimulation subjects would over-adjust such that confidence ratings would become higher than in the case of vertex stimulation.

It should be noted that all of the above arguments are indirect and therefore only suggestive of the fact that subjects used relatively constant decision criteria. Nevertheless, it is important to note that our VR model does not depend on the criteria being exactly the same between the low and high attention conditions. As long as the criteria for each condition are somewhat shifted in the direction of the criteria for the other condition, a process known as criterion attraction, the VR model would still predict similar results.

What are the implications of subjects' inability to use two separate sets of criteria? As we noted above, criterion setting likely depends on the prefrontal cortex (Lau and Rosenthal, 2012). Therefore, this limitation in criterion setting may be partially due to the limited processing resources of the prefrontal cortex. Another implication of the limitation in criterion setting is that our perceptual system cannot always perform optimal computations. Further experiments are needed to extend the current research (Gorea et al., 2005; Gorea and Sagi,

2000, 2001; Zak et al., in press) to high or low attention, or to different blocks of trials during brain stimulation.

*Variance reduction with attention*

Another key idea of the VR model is that attention reduces the trial-by-trial variability of an internal perceptual signal, which, in turn, reduces the probability with which the signal exceeds the decision criterion. This is a simple case of stochastic resonance (McDonnell and Abbott, 2009).

Direct evidence for this assumption comes from the fMRI study (Chapter 4), and more specifically from the analysis of the influence of pre-stimulus activity in the dorsal attention network (DAN) on the evoked activity in MT+. We found that high DAN activity leads to lower variability in the evoked signal in MT+, which suggests that high attention decreases the trial-by-trial variability of the perceptual signal.

This assumption is also supported by previous research. Psychophysics studies have shown that attention can exclude external noise (Dosher and Lu, 2000) as well as reduce internal noise (Lu and Dosher, 1998). Physiologically, attention is known to reduce the correlated noise among neuronal populations (Bressler and Silver, 2010; Cohen and Maunsell, 2009; Mitchell, Sundberg, and Reynolds, 2009), which should have the effect of reducing the trial-by-trial variability of the neuronal response when averaged over a large group of neurons. Critically, such population responses may be more important for conscious perception than single cell spiking activity (Maier et al., 2008). Although many single-neuron recording studies (Desimone and Duncan, 1995) and computational models (e.g., Reynolds and Heeger, 2009) focus on the effects of attention in boosting the perceptual signal, our findings highlight the importance of the effect of attention on reducing the variability of that signal. Also, it should be noted that what is critical here is trial-by-trial variability rather than noise exclusion

in a single trial. The variance reduction model implies that this noise reduction is a critical feature that explains the observed dissociation between accuracy and subjective ratings.

**Relationship between different kinds of subjective ratings**

In this work we have sometimes collected visibility ratings (Experiments 2.1-2.3 and metacognition experiment from Chapter 6), at other times confidence ratings (fMRI and TMS experiments from Chapters 4 and 5, respectively), and have even assessed certainty by asking subjects to wager points on their choices (Appendix B; Persaud, McLeod, and Cowey, 2007). These have been argued to be highly related measures (Merikle et al., 2001) though in some cases there may be subtle but important differences between them (Seth, 2008; Seth, Dienes, Cleeremans, Overgaard, and Pessoa, 2008). Nevertheless, for the purpose of this study we treat them as similar "subjective ratings." This is because according to the SDT framework, all of these ratings should behave in a similar fashion: regardless of the measure, high ratings should be given when the signal exceeds a certain threshold. Future experiments should verify this assumption by systematically varying the type of subjective ratings collected using the same kind of paradigm.

**Relation to the debates about attention and appearance, and attention and consciousness**

*Attention and appearance*

Recent research has explored the question of whether attention changes appearance (Carrasco et al., 2004; Fuller and Carrasco, 2006; Liu et al., 2006; Liu et al., 2009; Prinzmetal et al., 1997; Prinzmetal et al., 1998; Prinzmetal et al., 2008; Schneider and Komlos, 2008). On the surface, it may seem that the results of Carrasco et al. (2004) are contradictory to what we found here. Indeed, what we are arguing in this dissertation is that attention leads to

conservative subjective biases, while Carrasco et al. argued that attention leads to intensifying the subjective percept.

However, there are several large methodological differences between Carrasco et al.'s and our studies. First, Carrasco et al. asked subjects to judge the contrast of the stimulus, while we asked them to discriminate between a left and a right tilt, and give a subjective rating about the certainty in this decision. Second, Carrasco et al. asked subjects to make a comparative judgment, while we asked them to make a judgment on a single stimulus. These methodological differences mean that it is possible that Carrasco et al.'s and our task rely on different computational mechanisms. Indeed, it is not impossible for attention to make stimuli appear as having higher contrast while still decreasing subjects' confidence when they need to discriminate the identity of these stimuli.

In fact, the VR model could account for both our and Carrasco et al.'s results. According to the VR model, attention reduces variance while keeping the decision criteria constant. However, one additional feature of the model, that we have hitherto not focused on, is that attention also increases the signal strength. Indeed, when the VR model was implemented to fit the data from Chapters 1 and 2, we allowed attention to decrease variability of the SDT distributions while increasing the distance between their peaks.

Therefore, it is possible that the increased signal strength of the attended stimuli results in higher apparent contrast. This is in line with Carrasco et al.'s findings. On the other hand, as we explained previously (see Chapter 3), the increased variability of the uncued stimulus means that it would receive disproportionately high subjective ratings. This, in turn, is in line with our findings in this paper.

Thus, it appears that the task that subjects are asked to perform can lead to seemingly opposite results even if the underlying process is the same. It would be important to test this

hypothesis in a study that includes both kinds of questions (comparison between cued and uncued stimuli and confidence for a single stimulus) in the same subjects.

*Attention and consciousness*

Our results also relate to the debate of whether attention and consciousness are one and the same thing (Baars and Laureys, 2005; Dehaene et al., 2006; Frith, 2003; Koch and Tsuchiya, 2006; Rees et al., 2002; van Boxtel et al., 2010; Wilimzig et al., 2010). We found that attention led to lower subjective ratings, which is indicative of weaker conscious experience (Lau, 2007). Therefore, our findings show that not only are attention and consciousness potentially dissociable phenomena, they can sometimes have negative influence on each other. This does not necessarily imply that attention is not needed for a full conscious experience. Indeed, it is possible that the high subjective ratings given to unattended stimuli are a result of an inattentional inflation of subjective perception that is qualitatively different from the full conscious experience that is normally coupled with attention.

**Implications of the current research**

*Inattentional inflation of subjective perception*

This work has shown that attention leads to lower subjective ratings of visibility and confidence. There are two potential components of these effects. First, it could be that weakly attended stimuli produce unreasonably high levels of subjective visibility. This possibility would suggest that the visibility of the less processed unattended stimuli is inflated. The existence of such an effect was recently suggested by de Gardelle, Sackur and Kouider (2009) who showed that brief unattended non-letters may be perceived as normal letters.

A second possibility is that the strongly attended stimuli produce unreasonably low levels of subjective visibility. This interpretation would suggest that the brain artificially deflates the visibility of the better processed attended stimuli.

In our experiments we did not have a benchmark for what subjective ratings would have been optimal, and therefore could not arbitrate between the above possibilities. Nevertheless, the finding that attention increases metacognitive sensitivity (see Chapter 6) suggests that the subjective ratings for attended stimuli are set more optimally, and therefore it is likely that the unattended stimuli receive inflated subjective visibility. We call this phenomenon "inattentional inflation of subjective perception" (IISP).

*Change blindness blindness*

The current findings may help to explain subjects' overconfidence in their ability to perform well in inattentional and change blindness experiments (Mack and Rock, 1998; Simons and Chabris, 1999; Simons and Levin, 1997). Usually naïve subjects are extremely certain that they would have noticed a large change in the visual field and are therefore very surprised when it is revealed that they have completely missed an otherwise salient event to which they did not attend. In the case of change blindness, researchers have named this effect "change blindness blindness" (Beck, Levin, and Angelone, 2007). Our results suggest that unattended stimuli and events that are only partially processed in the brain might be accompanied by a disproportionately high sense of subjective visibility. We propose that exactly this heightened sense of subjective visibility leads to subjects' overconfidence in their ability to detect unattended objects or events, and therefore accounts for the "change blindness blindness" phenomenon.

*Uniform visual field*

It is known that our visual field is far from uniform with large portions of it being processed only in a very coarse manner (Abrams, Nizam, and Carrasco, 2012). For example, phenomena like change and inattentional blindness discussed above show that sometimes we can fail to process large and sometimes salient stimuli in our environment (Mack and Rock, 1998; Rees et al., 1999; Rock et al., 1992; Simons and Chabris, 1999; Simons and Levin, 1997). Also, peripheral vision in general has limited processing resolution and color sensitivity (Gordon and Abramov, 1977). In this context, given that the quality of information gathered from different parts of the scene in front of us is so vastly different, it is very surprising that we have a sense of uniformity of the visual field.

Our results may explain this puzzle. The interpretation would be that we do not "see" the unattended objects in the entire visual field, in the sense that our processing capacity for these objects may be poor. Instead, a biased perceptual decision making mechanism leads us to frequently say "yes" during object detection, even though the signal may be weak. The same mechanism leads to higher subjective ratings of confidence and visibility when it comes to discriminating the identities of the objects, although we may be poor at such discrimination. These tendencies may reflect an inflated sense of subjective phenomenology for the unattended periphery of the visual field.

*Philosophical debates*

Our results also relate to theories of conscious awareness in perception. There has been a recent debate among philosophers about whether our conscious visual experiences are rich or sparse (Block, 2007; de Gardelle, Sackur, and Kouider, 2009; Kouider, de Gardelle, Sackur, and Dupoux, 2010). Philosopher Ned Block (2007) argues that even though we are only capable of reporting the details of a few attended objects, this may simply indicate a

failure of reporting or access. Block calls this more limited type of consciousness "access consciousness." According to him, through careful introspection one may find that subjective experiences "overflow" what is being reported, despite the fact that these experiences may be too intricate to report in conjunction with one another. This experiential component of consciousness is called "phenomenal consciousness." Thus, Block's argument is that phenomenal consciousness overflows access consciousness. Like several other recent studies (de Gardelle et al., 2009; Kouider et al., 2010), our experiments challenge the plausibility of Block's argument. One's subjective impression of what is seen in unattended locations is likely to be inflated, because of the way inattention biases the subjective perceptual decision-making mechanisms. Thus, the intuitive account of phenomenal consciousness, as something that we experience but cannot necessarily report, may be due to inflation similar to a perceptual illusion rather than the existence of a different type of consciousness.

*Evolutionary advantage*

In Chapter 1 we found that attention leads to conservative detection biases. Why has the visual system developed a mechanism for detection that is consistently biased? One speculative possibility is that the tendency of the visual system to detect unattended stimuli more liberally may be evolutionarily adaptive. Indeed, a missed event in the periphery of our visual field might turn out to have dire consequences, while a false alarm only costs the effort to turn one's attention in the direction of the stimulus and examine it more fully. On the other hand, too many false alarms for attended stimuli might be costly to the organism and thus attended events might be optimally detected with a conservative criterion. Thus, despite the fact that having a visual system that sets detection criteria that are consistently biased is suboptimal in many situations, it could also be beneficial in other situations that may have been evolutionarily relevant.

*Attention not always beneficial*

Our results suggest that diverting attention from a stimulus does not always have negative consequences. Indeed, we showed that attention can lead to lower confidence and conservative detection criteria, which would sometimes lead to suboptimal performance (such as lower proportion of correct answers in Experiment 1.6 where the criterion for the uncued stimuli was closer to optimality). This finding adds to a set of studies from different areas within psychology on the potentially negative effects of attention in certain situations. For example, in an elegant experiment, Yeshurun and Carrasco (1998) showed that the lack of attention can improve perceptual performance in a textural discrimination task. The logic behind the experiment was that attention makes spatial resolution more fine-grained. The researchers designed a task in which one benefits from coarse-grained spatial discrimination, and therefore found that attention impaired subjects' performance. In a different experiment, Olivers and Nieuwenhuis (2005) showed that diverting attention by playing music improved subjects' detection performance in the attentional blink. Finally, Dijksterhuis, Bos, Nordgren, and van Baaren (2006) found that attending to one's choice is advantageous for simple choices but becomes disadvantageous for complex choices, such as choosing a car. The present work extends these previous studies by showing that attention can sometimes lead to suboptimal biases.

**Conclusion**

This dissertation demonstrates that attention leads to previously unexpected conservative subjective biases. These biases are well explained by our variance reduction model. The model is tested with a number of paradigms, and is able to account for puzzling dissociations between objective and subjective perception in a variety of contexts. These

findings have implications to current debates related to attention and subjective perception, and may shed light to surprising features of our day-to-day conscious experience.

## References

Abrahamyan, A., Clifford, C. W., Arabzadeh, E., & Harris, J. A. (2011). Improving visual sensitivity with subthreshold transcranial magnetic stimulation. *Journal of Neuroscience, 31*(9), 3290-3294.

Abrams, J., Nizam, A., & Carrasco, M. (2012). Isoeccentric locations are not equivalent: the extent of the vertical meridian asymmetry. *Vision Research, 52*(1), 70-78.

Amassian, V. E., Cracco, R. Q., Maccabee, P. J., Cracco, J. B., Rudell, A., & Eberle, L. (1989). Suppression of visual perception by magnetic coil stimulation of human occipital cortex. *Electroencephalography and Clinical Neurophysiology, 74*, 458-462.

Anderson, J. S., Druzgal, T. J., Lopez-Larson, M., Jeong. E, K,, Desai, K., & Yurgelun-Todd, D. (2011). Network anticorrelations, global regression, and phase-shifted soft tissue correction. *Human Brain Mapping, 32*, 919-934.

Anton-Erxleben, K., Abrams, J., & Carrasco, M. (in press). Equality judgments cannot distinguish between attention effects on appearance and criterion: a reply to Schneider (2011). *Journal of Vision*.

Baars, B. J., & Laureys, S. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. In S. Laureys (Ed.), *The Boundaries of Consciousness: Neurobiology and Neuropathology* (pp. 45-53). Amsterdam, The Netherlands: Elsevier.

Bahrami, B., Carmel, D., Walsh, V., Rees, G., & Lavie, N. (2008). Spatial attention can modulate unconscious orientation processing. *Perception, 37*(10), 1520-1528.

Bahrami, B., Lavie, N., & Rees, G. (2007). Attentional load modulates responses of human primary visual cortex to invisible stimuli. *Current Biology, 17*(6), 509-513.

Baldassi, S., Megna, N., & Burr, D. C. (2006). Visual clutter causes high-magnitude errors. *PLoS Biology, 4*(3), e56.

Beck, M. R., Levin, D. T., & Angelone, B. (2007). Change blindness blindness: Beliefs about the roles of intention and scene complexity in change detection. *Consciousness and Cognition, 16*(1), 31-51.

Blackmore, S. J., Brelstaff, G., Nelson, K., & Trościanko, T. (1995). Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception, 24*(9), 1075-1081.

Blaser, E., Sperling, G., & Lu, Z.-L. (1999). Measuring the amplification of attention. *Proceedings of the National Academy of Sciences of the United States of America, 96*(20), 11681-11686.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences, 30*, 481-499.

Boyer, J. L., Harrison, S., Ro, T. (2005). Unconscious processing of orientation and color without primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 16875-16879.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433-436.

Breitmeyer, B. G., Ro, T., & Ogmen, H. (2004). A comparison of masking by visual and transcranial magnetic stimulation: Implications for the study of conscious and unconscious processing. *Consciousness and Cognition, 13*, 829-843.

Bressler, D. W., & Silver, M. A. (2010). Spatial attention improves reliability of fMRI retinotopic mapping signals in occipital and parietal cortex. *Neuroimage, 53*, 526-533.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York, NY: Springer.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin and Review, 7*, 26-48.

Calhoun, V. D., Adali, T., Pearlson, G. D., & Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping, 14*, 140-151.

Carandini, M. (2004). Amplification of trial-to-trial response variability by neurons in visual cortex. *PLoS Biology, 2*, e264.

Carrasco, M. (2011). Visual attention: the past 25 years. *Vision Research, 51*, 1484-1525.

Carrasco, M., Fuller, S., & Ling, S. (2008). Transient attention does increase perceived contrast of suprathreshold stimuli: a reply to Prinzmetal, Long, and Leonhardt (2008). *Perception and Psychophysics, 70*(7), 1151-1164.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience, 7*(3), 308-313.

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R. et al. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature Neuroscience, 3*, 369-378.

Clarke, F. R., Birdsall, T. G., & Tanner, J. (1959). Two types of ROC curves and definitions of parameters. *The Journal of the Acoustical Society of America, 31*(5), 629-630.

Cleeremans, A., Timmermans, B., & Pasquali, A. (2007). Consciousness and metarepresentation: A computational sketch. *Neural Networks, 20*(9), 1032-1039.

Clifford, C. W. G., Arabzadeh, E., & Harris, J. A. (2008). Getting technical about awareness. *Trends in Cognitive Sciences, 12*(2), 54-58.

Cohen, M. R., & Maunsell, J. H. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience 12*, 1594-1600.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*, 201-215.

Corthout, E., Uttl, B., Walsh, V., Hallett, M. & Cowey, A. (1999). Timing of activity in early visual cortex as revealed by transcranial magnetic stimulation. *NeuroReport, 10*, 2631-2634.

Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neurosciece, 6*(2), 119-126.

Damoiseaux, J. S., Rombouts, S. A., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., & Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 13848-13853.

DeCarlo, L. T., & Cross, D. V. (1990). Sequential effects in magnitude scaling: models and theory. *Journal of Experimental Psychology: General, 119*, 375-396.

de Gardelle, V., Sackur, J., & Kouider, S. (2009). Perceptual illusions in brief visual presentations. *Consciousness and Cognition, 18*(3), 569-577.

Dehaene, S., Changeux, J.P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences, 10*(5), 204-211.

Dehaene, S. & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition, 79*, 1-37.

del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain, 132*, 2531-2540.

Desimone, R., & Duncan, J. (1995). Neural Mechanics of Selective Visual Attention. *Annual Review of Neuroscience, 18*, 193-222.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science, 311*, 1005-1007.

Dorfman, D. D., & Alf, E. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals--Rating-method data. *Journal of Mathematical Psychology*, *6*(3), 487-496.

Dosher, B. A., & Lu, Z.-L. (2000). Noise exclusion in spatial cuing of attention. *Psychological Science, 11*, 139-146.

Downing, C. J. (1988). Expectancy and visual-spatial attention: effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance, 14*(2), 188-202.

Eichele, T., Debener, S., Calhoun, V. D., Specht, K., Engel, A. K., Hugdahl, K., von Cramon, D. Y., & Ullsperger, M. (2008). Prediction of human errors by maladaptive changes in event-related brain networks. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 6173-6178.

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. *Consciousness and Cognition, 9*, 288-307.

Fleet, M. L., Brigham, J. C, & Bothwell, R. K. (1987). The confidence-accuracy relationship: The effects of confidence assessment and choosing. *Journal of Applied Social Psychology, 17,* 171-187.

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision-making. *Journal of Neuroscience, 32*, 6117-6125.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science, 329*, 1541-1543.

Fox, M. D., Halko, M. A., Eldaief, M. C., & Pascual-Leone, A. (in press). Measuring and manipulating brain connectivity with resting state functional connectivity magnetic resonance imaging (fcMRI) and transcranial magnetic stimulation (TMS). *Neuroimage.*

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 9673-9678.

Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 10046-10051.

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience, 8*, 700-711.

Fox, M. D., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2007). Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. *Neuron, 56*, 171-184.

Fox, M. D., Zhang, D., Snyder, A. Z., & Raichle, M. E. (2009) The global signal and observed anticorrelated resting state brain networks. *Journal of Neurophysiology, 101*, 3270-3283.

Franca, M., Koch, G., Mochizuki, H., Huang, Y.-Z., & Rothwell, J. C. (2006). Effects of theta burst stimulation protocols on phosphene threshold. *Clinical Neurophysiology, 117*, 1808-1813.

Frith, C. (2002). Attention to action and awareness of other minds. Consciousness and *Cognition, 11*(4), 481-487.

Fuller, S., & Carrasco, M. (2006). Exogenous attention and color perception: Performance and appearance of saturation and hue. *Vision Research, 46*(23), 4032-4047.

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychonomic Bulletin and Review, 10*(4), 843-876.

Gobell, J., & Carrasco, M. (2005). Attention alters the appearance of spatial frequency and gap size. *Psychological Science, 16*, 644-651.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30*, 535-574.

Gordon, J., & Abramov, I. (1977). Color vision in the peripheral retina. II. Hue and saturation. *Journal of the Optical Society of America, 67*, 202-207.

Gorea, A., Caetta, F., & Sagi, D. (2005). Criteria interactions across visual attributes. *Vision Research, 45*(19), 2523-2532.

Gorea, A., & Sagi, D. (2000). Failure to handle more than one internal representation in visual detection tasks. *Proceedings of the National Academy of Sciences of the United States of America, 97*(22), 12380-12384.

Gorea, A., & Sagi, D. (2001). Disentangling signal from noise in visual contrast discrimination. *Nature Neuroscience, 4*(11), 1146-1150.

Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics*. New York: John Wiley & Sons, Inc.

Gur, M., Beylin, A., & Snodderly, D. M. (1997). Response variability of neurons in primary visual cortex (V1) of alert monkeys. *Journal of Neuroscience, 17*, 2914-2920.

Haligan, P. W., Fink, G. R., Marshall, J. C., & Vallar, G. (2003). Spatial cognition: evidence from visual neglect. *Trends in Cognitive Sciences, 7(3)*, 125-133.

Harris, J. A., Clifford, C. W., & Miniussi, C. (2008). The functional effect of transcranial magnetic stimulation: signal suppression or neural noise generation? *Journal of Cognitive Neuroscience, 20*(4), 734-740.

Harris, K. D., & Thiele, A. (2011). Cortical state and attention. *Nature Reviews Neuroscience, 12,* 509-523.

Helmholtz, H. von. (1866). *Treatise on physiological optics* (J.P.C. Southall, Ed.; Vols. 2 and 3). Rochester, NY: Optic Society of America.

Hesselmann, G., Kell, C. A., Eger, E., & Kleinschmidt, A. (2008). Spontaneous local variations in ongoing neural activity bias perceptual decisions. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 10984-10989.

Hesselmann, G., Kell, C. A., & Kleinschmidt, A. (2008). Ongoing activity fluctuations in hMT+ bias the perception of coherent visual motion. *Journal of Neuroscience, 28*, 14481-14485.

Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron, 45*, 201-206.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513-541.

James, W. (1983). *The principles of psychology*. Cambridge, MA: Harvard University Press. (Original work published 1890)

Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 92-104.

Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences of the United States of America, 103*(45), 17048-17052.

Jolij, J., & Lamme, V. A. F. (2005). Repression of unconscious information by conscious processing: evidence from affective blindsight induced by transcranial magnetic stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(30), 10747-10751.

Kammer, T. (2007). Masking visual stimuli by transcranial magnetic stimulation. *Psychological Research, 71*, 659-666.

Kastner, S., Demmer, I., & Ziemann, U. (1998). Transient visual field defects induced by transcranial magnetic stimulation over human occiptal lobe. *Experimental brain research, 118*, 19-26.

Kentridge, R. W., Heywood, C. A., & Weiskrantz, L. (2004). Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia, 42*(6), 831-835.

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science, 324*, 759-764.

Kim, C. Y., & Blake, R. (2005). Psychophysical magic: rendering the visible 'invisible'. *Trends in Cognitive Sciences, 9*, 381-388.

Kirkpatrick, S., Gelatt, C. D., Jr, & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671-680.

Koch, C., & Tsuchiya, N. (2006). Attention and consciousness: Two distinct brain processes. *Trends in cognitive sciences, 11*(1), 16-22.

Koivitso, M., Mantyla, T., & Silvanto, J. (2010). The role of early visual cortex (V1/V2) in conscious and unconscious visual perception. *Neuroimage, 51*, 828-834.

Koivitso, M., Railo, H., & Salminen-Vaparanta, N. (2011). Transcranial magnetic stimulation of early visual cortex interferes with subjective visual awareness and objective forced-choice performance. *Consciousness and Cognition, 20*(2), 288-298.

Kolb, F. C., & Braun, J. (1995). Blindsight in normal observers. *Nature, 377*(6547), 336-338.

Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science, 18*(1), 64-71.

Kouider, S., de Gardelle, V., Sackur, J., & Dupoux, E. (2010). How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences, 14*, 301-307.

Krantz, D. H. (1969). Threshold theories of signal detection. *Psychological Review, 76*, 308-324.

Lau, H. (2007). A higher order Bayesian decision theory of consciousness. *Progress in Brain Research, 168*, 35-48.

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 18763-18768.

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences, 15*(8), 365-373.

Laycock, R., Crewther, D. P., Fitzgerald, P. B., & Crewther, S. G. (2007) Evidence for fast signals and later processing in human V1/V2 and V5/MT+: A TMS study of motion perception. *Journal of Neurophysiology, 98*, 1253-1262.

Leh, S. E., Johansen-Berg, H., & Ptito, A. (2006). Unconscious vision: new insights into the neuronal correlate of blindsight using diffusion tractography. *Brain, 129*, 1822-1832.

Levin, D. T., Simons, D. J., Angelone, B. L., & Chabris, C. F. (2002). Memory for centrally attended changing objects in an incidental real-world change detection paradigm. *British Journal of Psychology, 93*, 289-302.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America, 99*(14), 9596-9601.

Liu, T., Abrams, J., & Carrasco, M. (2009). Voluntary attention enhances contrast appearance. *Psychological Science, 20*(3), 354-362.

Liu, T., Fuller, S., & Carrasco, M. (2006). Attention alters the appearance of motion coherence. *Psychonomic Bulletin and Review, 13*(6), 1091-1096.

Lu, Z.-H., & Dosher, B. A. (2008). Characterizing observers using external noise and observer models: Assessing internal representations with external noise. *Psychological Review, 115*, 44-82.

Lund, T. E., Norgaard, M. D., Rostrup, E., Rowe, J. B., & Paulson, O. B. (2005). Motion or activity: Their role in intra- and inter-subject variation in fMRI. *Neuroimage, 26*, 960-964.

Maccabee, P. J., Amassian, V. E., Cracco, R. Q., Cracco, J. B., Rudell, A. P., Eberle, L. P., & Zemon, V. (1991). Magnetic coil stimulation of human visual cortex: studies of perception. *Electroencephelogr Clin Neurophysiol 43*, 111-120.

Mack, A. & Rock, I. (1998). *Inattentional Blindness*. Cambridge, MA: The MIT Press.

Macmillan, N., & Creelman, C. (2005). *Detection theory: A user's guide*. New York, NY: Cambridge University Press.

Maier, A., Wilke, M., Aura, C., Zhu, C., Ye, F. Q., & Leopold, D. A. (2008). Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nature Neuroscience, 11*, 1193-2000.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition, 21*(1), 422-430.

McDonnell, M. D., & Abbott, D. (2009). What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology. *PLoS Computational Biology, 5*, e1000348.

Merikle, P. M., Smilek, D., & Eastwood, J. D. (2001). Perception without awareness: perspectives from cognitive psychology. *Cognition, 79*(1-2), 115-134.

Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: Knowing About Knowing*. Cambridge, MA: MIT Press.

Miller, B. M., Fendrich, R., Eliassen, J. C., Demirel, S., & Gazzaniga, M. S. (1996). Transcranial magnetic stimulation: delays in visual suppression due to luminance changes. *Neuroreport, 7*, 1740-1744.

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron, 55*, 131-141.

Montagna, B., & Carrasco, M. (2006). Transient covert attention and the perceived rate of flicker. *Journal of Vision, 6*(9), 955-965.

Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences*, *14*(10), 435-440.

Mori, S. (1998). Effects of stimulus information and number of stimuli on sequential dependencies in absolute identification. *Canadian Journal of Experimental Psychology, 52,* 72-83.

Mueller, S. T., & Weidemann, C. T. (2008). Decision noise: An explanation for observed violations of signal detection theory. *Psychonomic Bulletin and Review, 15*(3), 465-494.

Murphy, K., Birn, R. M., Handwerker, D. A., Jones, T. B., & Bandettini, P. A. (2009). The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *Neuroimage, 44*, 893-905.

Naccache, L., Blandin, E., & Dehaene, S. (2002). Unconscious masked priming depends on temporal attention. *Psychological Science, 13*(5), 416-424.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109-133.

Nikolov, S., Rahnev, D. A., & Lau, H. C. (2010). Probabilistic model of onset detection explains paradoxes in human time perception. *Frontiers in Psychology, 1*, 37.

Northoff, G., & Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in Cognitive Sciences, 8*, 102-107.

Olivers, C. N. L., & Nieuwenhuis, S. (2005). The Beneficial Effect of Concurrent Task-Irrelevant Mental Activity on Temporal Attention. *Psychological Science, 16*(4), 265-269.

Overgaard, M., Nielsen, J.F., & Fuglsang-Frederiksen, A. (2004). A TMS study of the ventral projections from V1 with implications for the finding of neural correlates of consciousness. *Brain and Cognition, 54*, 58-64.

Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology, 52*, 629-651.

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience, 10*(2), 257-261.

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review, 117*, 864-901.

Pollack, I., & Hsieh, R. (1969). Sampling variability of the area under the ROC-curve and of d'e. *Psychological Bulletin, 71*(3), 161-173.

Posner, M. I., & Peterson, S. E. (1990). The Attention System of the Human Brain. *Annual Review of Neuroscience, 13(1)*, 25-42.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology, 109*(2), 160-174.

Prigatano, G. P., & Schacter, D. L. (1991). *Awareness of deficit after brain injury: clinical and theoretical issues*. Oxford: Oxford University Press.

Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention: I. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance, 24*(1), 261-282.

Prinzmetal, W., Long, V., & Leonhardt, J. (2008). Involuntary attention and brightness contrast. *Perception and Psychophysics, 70*(7), 1139-1150.

Prinzmetal, W., Nwachuku, I., Bodanski, L., Blumenfeld, L., & Shimizu, N. (1997). The phenomenology of attention. 2. Brightness and contrast. *Consciousness and Cognition, 6*(2-3), 372-412.

Rahnev, D., Bahdo, L., de Lange, F. P., & Lau, H. (in press). Pre-Stimulus hemodynamic activity in dorsal attention network is negatively associated with decision confidence in visual perception. *Journal of Neurophysiology*.

Rahnev, D., Huang, E., & Lau, H. (2012). Subliminal stimuli in the near absence of attention influence top-down cognitive control. *Attention, Perception, and Psychophysics, 74*(3), 521-532.

Rahnev, D. A., Kok, P., Munneke, M., Bahdo, L., de Lange, F. P., & Lau, H. (in prep.). The neural mechanisms of occipital theta-burst transcranial magnetic stimulation.

Rahnev, D. A., & Lau, H. (in prep.). Attention improves metacognitive sensitivity.

Rahnev, D., Lau, H., & de Lange, F. P. (2011). Prior expectation modulates the interaction between sensory and prefrontal regions in the human brain. *Journal of Neuroscience 31*(29), 10741-10748.

Rahnev, D. A., Maniscalco, B., Luber, B, Lau, H., & Lisanby, S. H. (2012). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *Journal of Neurophysiology, 107*(6), 1556-1563.

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F. P., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience, 14*(12), 1513-1515.

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation, 20*, 873-922.

Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(4)*, 763-785.

Rees, G., Kreiman, G., & Koch, C. (2002). Neural correlates of consciousness in humans. *Nature Reviews Neuroscience, 3*(4), 261-270.

Rees, G., Russell, C., Frith, C. D., & Driver, J. (1999). Inattentional blindness versus inattentional amnesia for fixated but ignored words. *Science, 286*, 2504-2507.

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron, 61(2)*, 168-185.

Ro, T., Breitmeyer, B., Burton, P., Singhal, N. S., & Lane, D. (2003). Feedback contributions to visual awareness in human occipital cortex. *Current Biology, 13*(12), 1038-1041.

Ro, T., Shelton, D. J. M., Lee, O. L., & Chang, E. (2004). Extrageniculate mediation of unconscious vision in transcranial magnetic stimulation induced blindsight. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 9933-9935.

Rock, I., Linnett, C. M., Grant, P., & Mack, A. (1992). Perception without attention: results of a new method. Cognitive Psychology, 24, 502-534.

Roos, K. L., Tuite, P. J., Below, M. E., & Pascuzzi, R. M. (1990). Reversible cortical blindness (Anton's syndrome) associated with bilateral occipital EEG abnormalities. *Clinical Electroencephalography, 21*(2), 104-109.

Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. C. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual. *Cognitive Neuroscience, 1*, 1-11.

Rudin, D. O., & Eisenman, G. (1954). The action potential of spinal axon in vitro. *Journal of General Physiology, 37*, 505-538.

Rushton, W. A. (1927). Effect upon the threshold for nervous excitation of the length of nerve exposed and the angle between current and nerve. *Journal of Physiology, 63*, 357-377.

Ruzzoli, M., Abrahamyan, A., Clifford, C. W. G., Marzi, C. A., Miniussi, C., and Harris, J. A. (2011). The effects of TMS on visual motion sensitivity: an increase in neural noise or a decrease in signal strength? *Journal of Neurophysiology, 106*, 138-143.

Ruzzoli, M., Marzi, C. A., & Miniussi, C. (2010). The neural mechanisms of the effects of transcranial magnetic stimulation on perception. *Journal of Neurophysiology, 103*(6), 2982-2989.

Sadaghiani, S., Hesselmann, G., & Kleinschmidt, A. (2009). Distributed and antagonistic contributions of ongoing activity fluctuations to auditory stimulus detection. *Journal of Neuroscience, 29(42)*, 13410-13417.

Sadaghiani, S., Scheeringa, R., Lehongre, K., Morillon, B., Giraud, A. L., & Kleinschmidt, A. (2010). Intrinsic connectivity networks, alpha oscillations, and tonic alertness: a simultaneous electroencephalography/functional magnetic resonance imaging study. *Journal of Neuroscience, 30*, 10243-10250.

Salminen-Vaparanta, N., Noreika, V., Revonsuo, A., Koivisto, M., & Vanni, S. (2012). Is selective primary visual cortex stimulation achievable with TMS? *Human Brain Mapping, 33*, 652-665.

Sapir, A., d'Avossa, G., McAvoy, M., Shulman, G. L., & Corbetta, M. (2005). Brain signals for spatial attention predict performance in a motion discrimination task. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 17810-17815.

Schmid, M. C., Mrowka, S. W., Turchi, J., Saunders, R. C., Wilke, M., Peters, A. J., Ye, F. Q., & Leopold, D. A. (2010). Blindsight depends on the lateral geniculate nucleus. *Nature, 466*, 373-377.

Schneider, K. A., & Komlos, M. (2008). Attention biases decisions but does not alter appearance. *Journal of Vision, 8*(15), 1-10.

Schwarzkopf, D. S., Silvanto, J., & Rees, G. (2011). Stochastic resonance effects reveal the neural mechanisms of transcranial magnetic stimulation. *Journal of Neuroscience, 31*(9), 3143-3147.

Seth, A. K. (2008). Post-decision wagering measures metacognitive content, not sensory consciousness. *Consciousness and Cognition, 17*(3), 981-983.

Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends in Cognitive Sciences, 12*, 314-321.

Siebner, H. R., Hartwigsen, G., Kassuba, T., & Rothwell, J. C. (2009). How does transcranial magnetic stimulation modify neuronal activity in the brain? Implications for studies of cognition. *Cortex, 45*(9), 1035-1042.

Silvanto, J., Cowey, A., Lavie, N., & Walsh, V. (2005). Striate cortex (V1) activity gates awareness of motion. *Nature Neuroscience, 8*, 143-144.

Silvanto, J., Lavie, N., & Walsh, V. (2005). Double dissociation of V1 and V5/MT activity in visual awareness. *Cerebral cortex, 15*, 1736-1741.

Silvanto. J, & Muggleton, N. G. (2008). New light through old windows: moving beyond the "virtual lesion" approach to transcranial magnetic stimulation. *Neuroimage, 39*, 549-552.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, *28*, 1059-1074

Simons, D. J., & Levin, D. T. (1997). Change Blindness. *Trends in Cognitive Sciences, 1*(7), 261-267.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*, 1-29.

Szczepanowski, R., & Pessoa, L. (2007). Fear perception: can objective and subjective awareness measures be dissociated? *Journal of Vision, 7*, 1-17.

Tark, K. J., & Curtis, C. E. (2009). Persistent neural activity in the human frontal cortex when maintaining space that is off the map. *Nature Neuroscience, 12*, 1463-1468.

Thielscher, A., Reichenbach, A., Uğurbil, K., & Uludağ, K. (2010). The cortical site of visual suppression by transcranial magnetic stimulation. *Cerebral Cortex, 20*(2), 328-338.

Tsal, Y., Shalev, L., Zakay, D., & Lubow, R. E. (1994). Attention reduces perceived brightness contrast. *The Quarterly journal of experimental psychology, 47*(4), 865-893.

van Boxtel, J. J., Tsuchiya, N., & Koch, C. (2010). Opposing effects of attention and consciousness on afterimages. *Proceedings of the National Academy of Sciences of the United States of America, 107*(19), 8883-8888.

Ward, L. M. (1979). Stimulus information and sequential dependencies in magnitude estimation and cross-modality matching. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 444-459.

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception and Psychophysics, 33(2)*, 113-120.

Weiskrantz, L. (1986). *Blindsight: A Case Study and Implications*. Oxford: Oxford University Press.

Wilimzig, C., Tsuchiya, N., Fahle, M., Einhäuser, W., & Koch, C. (2008). Spatial attention increases performance but not subjective confidence in a discrimination task. *Journal of Vision, 8*(5), 1-10.

Wyart, V., & Tallon-Baudry, C. (2008). A neural dissociation between visual awareness and spatial attention. Journal of Neuroscience, 28, 2667-2679.

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature, 396*, 72-75.

Zak, I., Katkov, M., Gorea, A., & Sagi, D. (in press). Decision criteria in dual discrimination tasks estimated using external-noise methods. *Attention, Perception, and Psychophysics*.

## Figure and Table Captions

**Figure 1**. Task design in Experiments 1.1-1.6. Subjects were instructed to fixate on a central red square throughout the duration of the experiment (shown here in black). Each trial began with a 500 ms presentation of a pre-cue. Stimuli consisted of noise background composed of uniformly distributed intensity values with a superimposed grating of varying contrast, and were presented for 367 ms (except for Experiment 1.4 where the stimuli were presented for 50 ms). The gratings had random orientations and subjects needed to indicate whether gratings appeared in the diagonal post-cued after the presentation of the stimuli. On 70% of the trials the post-cue coincided with the pre-cue ("cued trials") while on the remaining 30% it did not ("uncued trials").

**Figure 2.** Capacity d' and criterion c in Experiment 1.1. We matched the detection sensitivity (d') between the cued and uncued stimuli, and found that cued stimuli had higher (more conservative) criterion compared to uncued stimuli. Here and in the rest of the figures, data are means ± s.e.m.

**Figure 3.** The influence of the unprobed stimulus on the criterion c in Experiment 1.1. To investigate the putative role of the stimulus in the "unprobed" location (i.e., the stimulus that was not response-cued), we plotted hit and false alarm rates separately for the cases when the unprobed location contained a target or a non-target. The figure shows that hit and false alarm rates were higher in the uncued condition regardless of the identity of the unprobed location.

**Figure 4.** Capacity d' and criterion c in Experiments 1.2-1.5. To investigate the robustness of the results in Experiment 1.1, we replicated the experiment four additional times while

changing different aspects of the design (see main text in Chapter 1). As in Experiment 1.1, in all four experiments we titrated the contrast of the gratings so that subjects performed equally well in the cued and uncued trials and observed significant differences between the cued and uncued conditions in criterion but not in d'.

**Figure 5.** Eye-tracking results from Experiment 1.5. (**a**) We plotted the average horizontal and vertical eye position during the stimulus presentation for each of the 960 trials for one representative subject who showed variability closest to the group average (i.e., not the "best" subject). The standard deviation of the fixation positions for this subject was 0.15 degrees for the horizontal dimension and 0.67 for the vertical dimension. (The average standard deviations across the 6 subjects were 0.23 and 0.55 for the horizontal and vertical dimensions, respectively.) (**b**) We show the fixations for one trial from the subject from (a).

**Figure 6**. Relationship between d' and criterion c in Experiment 1.6. Each point represents one contrast level for one subjects. Least-squares linear fits were performed for each subject separately. The lines displayed in the figure are the means of the individual fits.

**Figure 7**. Relationship between contrast level and criterion c in Experiment 1.6. Mean criterion c across the nine subjects was plotted for each contrast value separately for cued and uncued gratings. Lines represent the least-squares linear fits, averaged across subjects.

**Figure 8.** Hit and false alarm rates in Experiment 1.6. We plotted the hit and false alarm rates for each contrast level for the cued and uncued trials separately. The figure shows that even though hit rate increased with contrast, false alarm rate stayed relatively constant across the different contrast levels.

**Figure 9.** Task design in Experiments 2.1 and 2.2. Subjects were instructed to fixate on a central red square throughout the duration of the experiment (shown here in black). Each trial began with a 500 ms presentation of a pre-cue. Stimuli consisted of noise background composed of uniformly distributed intensity values with a superimposed grating of varying contrast, and were presented for 367 ms. Only gratings tilted to 45° and 135° were used and subjects needed to indicate whether the tilt of the gratings in the post-cued diagonal was 45° or 135° and report their visibility. On 70% of the trials the post-cue coincided with the pre-cue ("cued trials") while on the remaining 30% it did not ("uncued trials").

**Figure 10.** Capacity d' and visibility ratings in Experiment 2.1. We matched the detection sensitivity (d') between the cued and uncued stimuli, and found that cued stimuli led to lower subjective ratings of visibility compared to uncued stimuli.

**Figure 11**. Relationship between d' and visibility ratings in Experiment 2.2. Each point represents one contrast level for one subject. Least-squares linear fits were performed for each subject separately. The lines displayed in the figure are the means of the individual fits.

**Figure 12**. Relationship between contrast and visibility ratings in Experiment 2.2. Mean visibility rating across all subjects was plotted for each contrast value separately for cued and uncued gratings. Lines represent the least-squares linear fits, averaged across subjects.

**Figure 13.** Task design in Experiment 2.3. We presented different numbers of stimuli on the screen in order to manipulate how subjects distribute their attention to different objects. In one condition we used 2 items on the screen (a relatively focused mode of attention), and in the

other we used 4 items on the screen (a relatively distributed mode of attention). Subjects had to indicate the tilt (left/right) of a post-cued Gabor patch and rate the visibility (high/low) of the tilt of that patch.

**Figure 14.** Capacity d' and visibility ratings in Experiment 2.3. Both d' and visibility ratings increased with higher contrast. Further, d' was roughly similar for the 2-patch task with 6% contrast, and the 4-patch task with 8%, 10%, and 12% contrast (see the horizontal dashed line). Nevertheless, compared to the 6%-contrast 2-patch task, the visibility of the grating was judged to be higher for the 4-patch task for the 8% contrast, 10% contrast, and 12% contrast.

**Figure 15.** A depiction of discrimination between two classes of stimuli within the framework of signal detection theory According to signal detection theory (SDT), left- and right-tilted bars produce Gaussian distributions of internal signals along a single dimension. A discrimination criterion is used to distinguish between these two classes of stimuli. Confidence or visibility responses are given based on additional confidence criteria. The figure represents a situation in which subjects needed to place two such criteria. The four types of responses would then naturally fall between consecutive confidence and discrimination criteria. Further, high attention may decrease the noise in the distributions (upper part of the figure) compared to low attention (lower part of the figure). Thus, an increase in attention can lead to an increase in d' and a simultaneous decrease in the number of trials judged with high confidence or visibility.

**Figure 16.** Data fits of the variance reduction (VR) model. Fits are presented for (a) Experiment 1.1, (b) Experiment 1.6, (c) Experiment 2.1, and (d) Experiment 2.2. In (b) and (d) we marked the contrasts with the numbers 1 to 4 where 1 stands for the lowest presented

contrast and 4 stands for the highest contrast. The observed data for these experiments are presented in Figures 2, 7, 10, and 12, respectively.

**Figure 17.** A depiction of the VR model. The VR model was implemented for (a) Experiment 1.1, (b) Experiment 1.6, (c) Experiment 2.1, and (d) Experiment 2.2. The model postulates that for all experiments attention leads to smaller trial-by-trial variability in the internal perceptual signal but does not change the decision criteria used.

**Figure 18.** Models in which attention does not affect the variability of the distributions cannot account for the observed data. (a) A depiction of the standard 'mean-only' model in which the Target-Absent distribution is identical for the cued and uncued trials. (b) A depiction of an alternative version of the 'mean-only' model which can mathematically account for the observed results but which necessitates that attention shifts the Target-Absent distribution to the right. See main text for details.

**Figure 19.** Task design in fMRI experiment (Chapter 4). We employed a sparse design with inter-trial intervals (ITI) of 20 to 40 seconds to allow for the BOLD signal to return to baseline. During that period subjects viewed stationary dots. Trials consisted of 100 ms coherent dot motion that was followed by a "mask" of random motion (400 ms). Subjects were asked to judge the motion direction (expanding / contracting) and then provide a confidence rating on a 1-4 scale.

**Figure 20.** Localization of the dorsal attention network (DAN). DAN was identified using standard seed-based methods (see Methods in Chapter 4). The network included the frontal eye fields and intraparietal regions.

**Figure 21.** Average time courses in DAN. The panel on the left plots the timecourse of high and low confidence trials, while the panel of the right plots the timecourse of correct and error trials. The vertical dashed line shows the average time of the second button press (the confidence response).

**Figure 22.** Mean pre-stimulus signal for each of the four high/low attention x correct/error conditions. Mean pre-stimulus signal was defined as the average BOLD activity in the two scans prior to stimulus onset.

**Figure 23.** Evoked activations in MT+. (a) The medial temporal cortex (MT/V5) was identified for each subject using the seed-based analysis used to identify the attentional network. (b) Compared to low confidence trials, high confidence trials resulted in higher evoked activity in MT+.

**Figure 24.** Results in the default network. (a) Pre-stimulus activity in the default mode network. The network was identified using a seed-based analysis with a seed in the posterior cingulate (PCC). The network matches pretty well what previous research has found: it included PCC and bilateral parietal cortex, although the medial prefrontal cortex (MPF) was less prominent than in earlier work. (b) The time courses for low/high confidence trials as well as for correct/error trials show negative evoked activations, which is not surprising given the fact that the default network is suppressed during demanding tasks. Conversely to the attentional network, the pre-stimulus activity was higher for high confidence than for low confidence trials. The vertical dashed line shows the average time of the second button press

(the confidence response). (c) We separated the trials by confidence and accuracy and plotted the pre-stimulus activity.

**Figure 25.** Results in the alertness network. (a) The network was identified using a seed in the dorsal ACC (dACC; see Methods in Chapter 4) and corresponded well with previous research. It included dorsal dACC, bilateral anterior insula and anterior thalamus. (b) We extracted time courses for low and high confidence trials, as well as for correct and error trials. The vertical dashed line shows the average time of the second button press (the confidence response). (c) We separated the trials by confidence and accuracy and plotted the pre-stimulus activity.

**Figure 26.** Attention decreases the variability of the evoked response in MT+. The figure shows the Fano factor (a measure of variability, see Methods in Chapter 4) for the distributions of evoked responses of MT+ for the trials that had either low or high pre-stimulus dorsal attention network activity. DAN = dorsal attention network.

**Figure 27.** Task design in the TMS experiment (Chapter 5). The stimulus was a small white bar presented below fixation (exact location was chosen for each subject according to his or her phosphene location). The TMS pulse was delivered 100 ms after the onset of the stimulus. Throughout the experiment subjects were asked to fixate on a small white cross that changed color to black when the stimulus appeared. Subjects gave their response with a single button press indicating both the tilt of the bar (left/right) and their confidence (high/low). The duration of each trial was exactly 5 seconds.

**Figure 28.** Effect of occipital TMS on capacity d'. TMS to the occipital cortex decreased d' for each of the six subjects. The group means are plotted on the right side of the figure.

**Figure 29.** Effect of occipital TMS on confidence ratings. Unlike capacity d', confidence ratings increased after occipital TMS. The effect was present for five of the six subjects. The group means are plotted in the right side of the figure.

**Figure 30.** A depiction of the five models used in the TMS experiment. The first three models postulated a single processing channel in which occipital TMS is posited to affect the signal only (Model 1), the noise only (Model 2), or the signal and the noise (Model 3), respectively. The other two models were dual-channel models that included 'conscious' and 'unconscious' channels that operate in parallel, with confidence rating being specifically dependent on the 'conscious' channel. In Model 4 occipital TMS affected the signal of the 'conscious' channel only, while in Model 5 occipital TMS affected the signal in both the 'conscious' and 'unconscious' channels.

**Figure 31.** Model selection results. Formal model comparison was conducted using Akaike Information Criterion (AIC), which rewards models for closely fitting observed data while punishing them for the degree of complexity. Higher average Akaike weight indicates that the corresponding model is more likely. Models 1-5 are displayed in order from left to right and are indexed by a short description indicating whether the corresponding model postulates one or two processing channels, and what parameters are affected by occipital TMS (C channel, 'conscious' channel; C and U channels, 'conscious' and 'unconscious' channels; for more detail on each model see Methods in Chapter 5). Model 2 – which assumes that TMS only influenced the variability of the visual signal within a single process but not the average signal strength – was the clear winner with average Akaike weights about two times higher than the

second best model. A complementary analysis using the Bayesian Information Criterion (BIC) produced similar results with Model 2 winning over the rest of the models.

**Figure 32.** Fitted data of the winning Model 2. (a) Model fits for each individual subject for the signal detection measure d'. Group means are plotted on the right as in Figure 28. (b) Model fits for each individual subject for the confidence ratings. Group means are plotted on the right as in Figure 29.

**Figure 33.** Performance results in the experiment on metacognition (Chapter 6). We plotted the individual data for d' (left panel), visibility ratings (middle panel), and reaction time (RT; right panel). As can be seen from the figure, d' values were roughly equated between the cued and uncued conditions. However, visibility ratings and reaction times were consistently higher for the uncued stimuli. The diagonal line in each panel represents the area where cued and uncued values are equal.

**Figure 34.** Results on metacognitive sensitivity. We plotted the individual data for three different measures of metacognitive sensitivity: meta-d' (left panel), $A_{roc}$ (middle panel), and phi (right panel). As can be seen from the figure, for all measures, metacognitive sensitivity was higher for the cued stimuli compared to the uncued stimuli. The diagonal line in each panel represents the area where cued and uncued values are equal.

**Figure 35.** Model fits of the VCJR model for performance and metacognition. We plotted the model fits for each subject for d' (left panel), visibility ratings (middle panel), and $A_{roc}$ (right panel). As can be seen from the figure, the model provided good fits to the data (compare with

observed values in Figures 33 and 34). The diagonal line in each panel represents the area where cued and uncued values are equal.

**Table 1.** Pre-stimulus activity in some common brain networks. We computed the average pre-stimulus activity for each trial type (see Figure 22) in different brain networks. The dorsal attention network was also defined in two different ways using either seed-based methods or independent component analysis (ICA, see Methods in Chapter 4). One subject was excluded from this statistical analysis because he did not have any high confidence error trials.

**Table 2.** Fit from our computational simulations. We carried out simple computer simulations (see Methods in Chapter 4) to investigate if the VR model can fit the observed data from the experiment. The fitting procedure was performed four different times to ensure that the fitting procedure did not get "stuck" in local minima. This table reports the fitted values produced by the first fit which provided the best match to the observed values. The values were generated by simulating 15,000,000 trials.

## Figures and Tables

**Figure 1. Task design in Experiments 1.1-1.6**

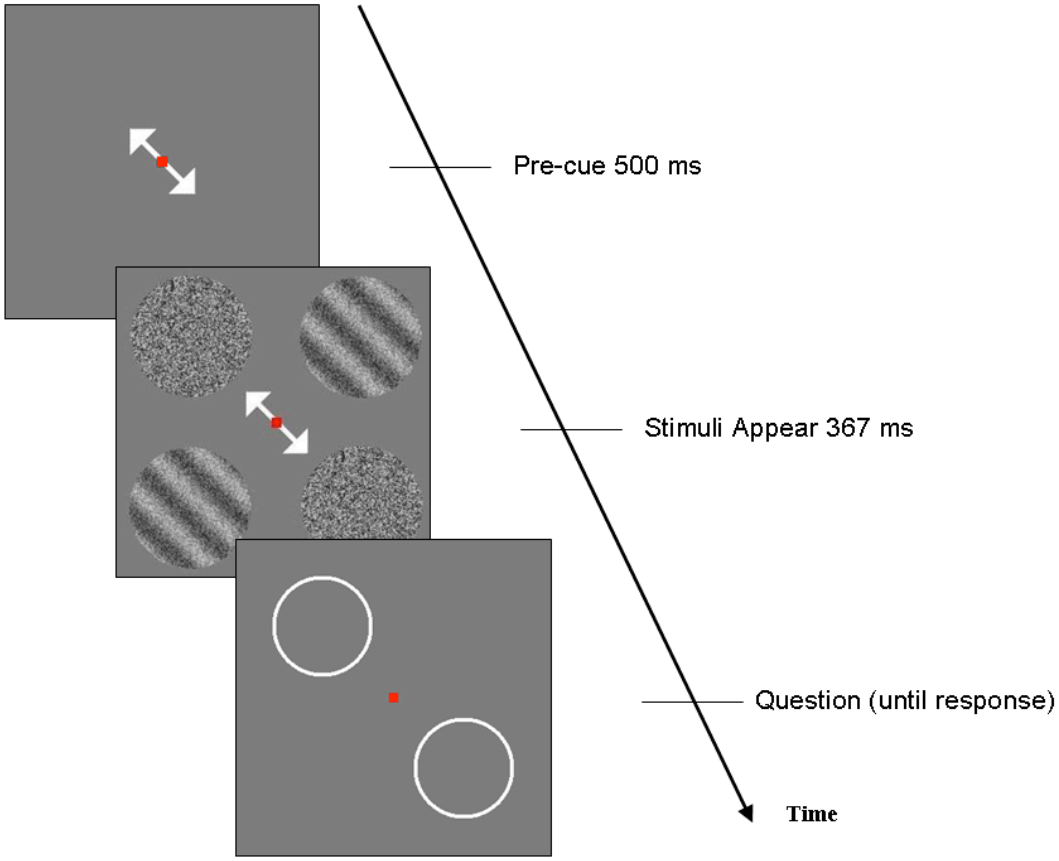

**Figure 2. Capacity d' and criterion c in Experiment 1.1**

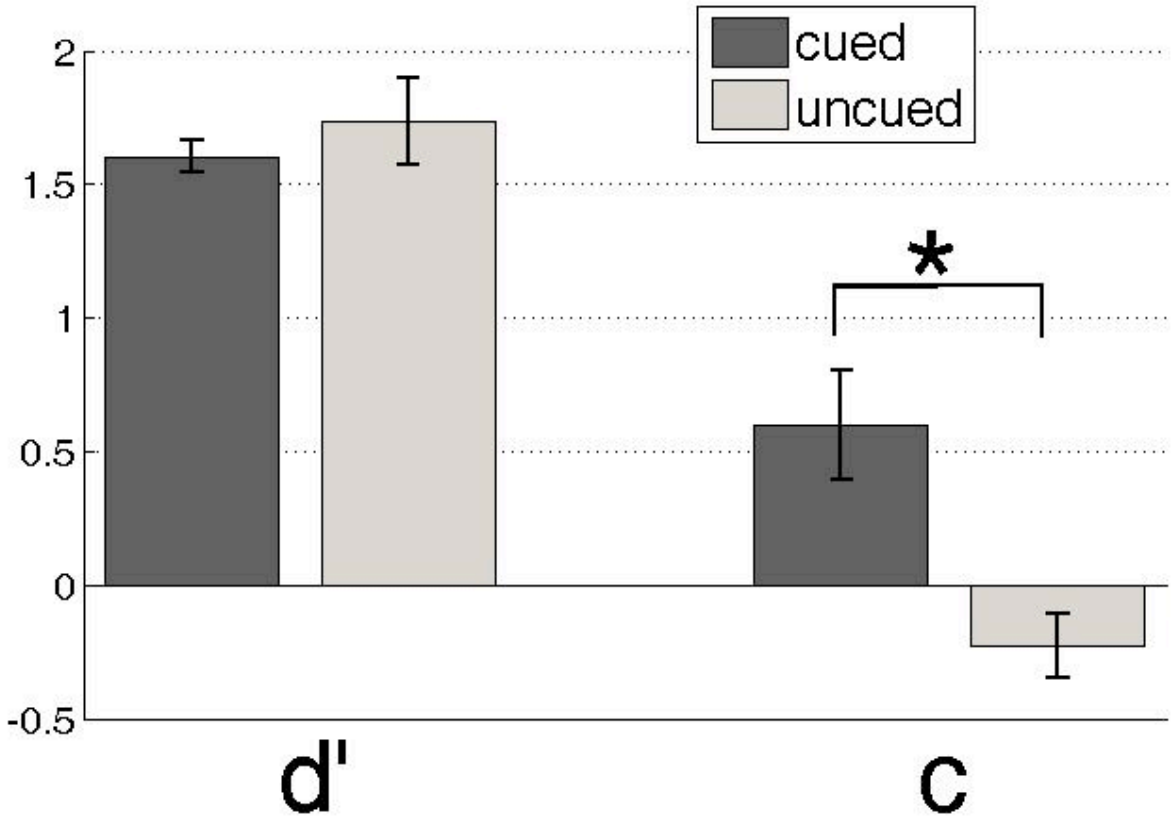

**Figure 3. The influence of the unprobed stimulus on the criterion c in Experiment 1.1**

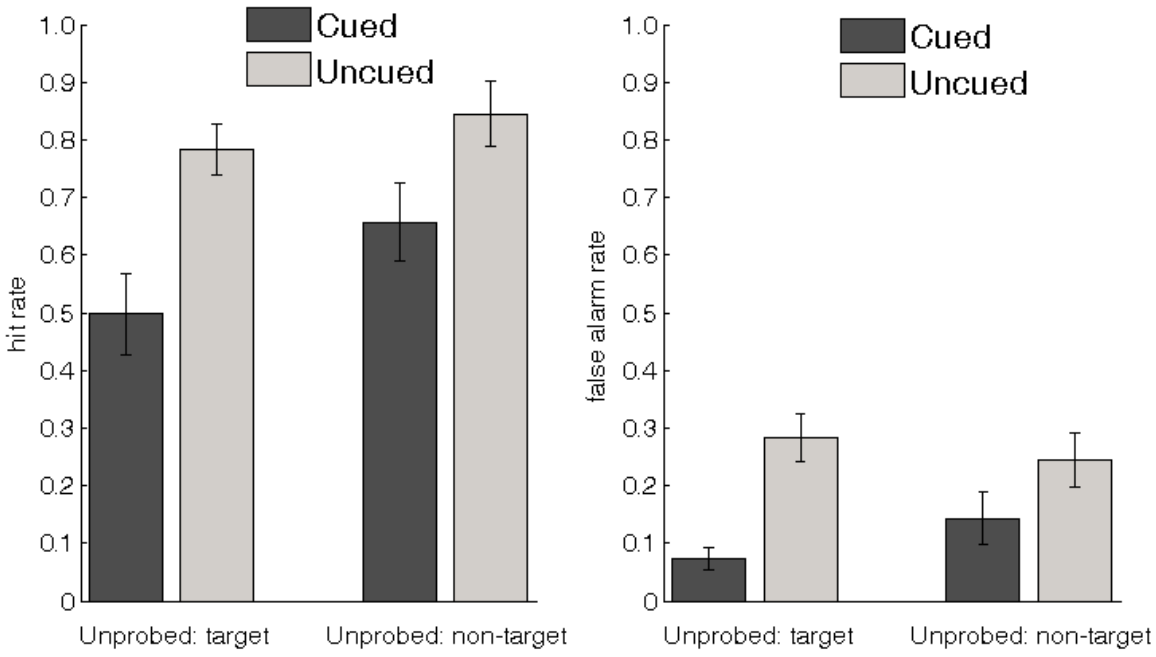

**Figure 4. Capacity d' and criterion c in Experiments 1.2-1.5**

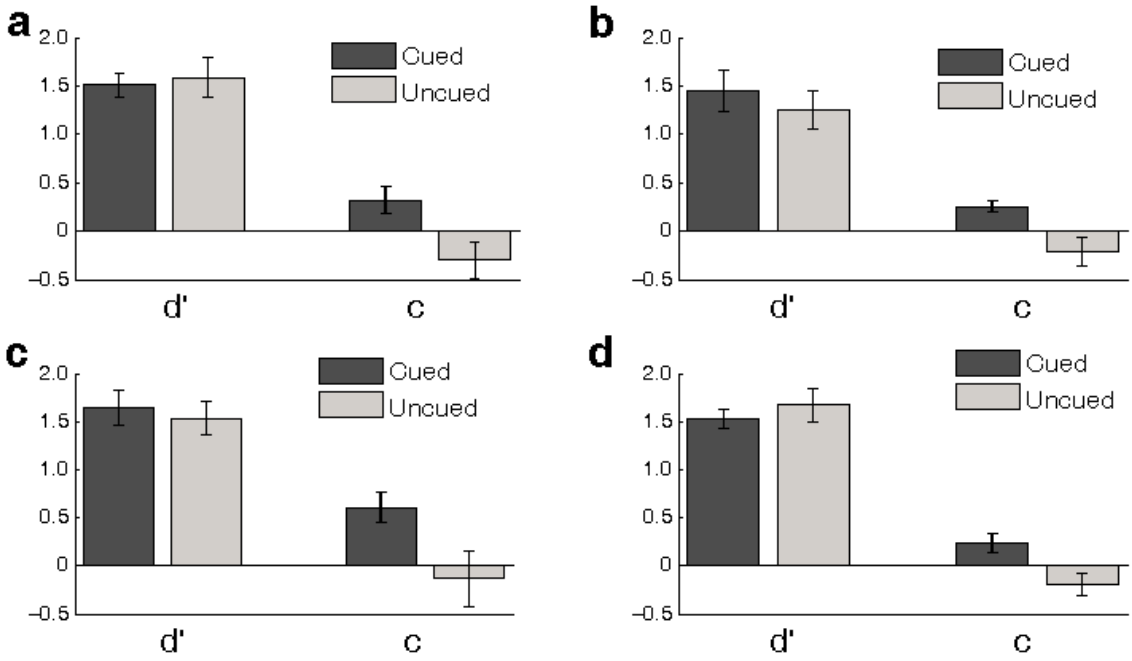

**Figure 5. Eye-tracking results from Experiment 1.5**

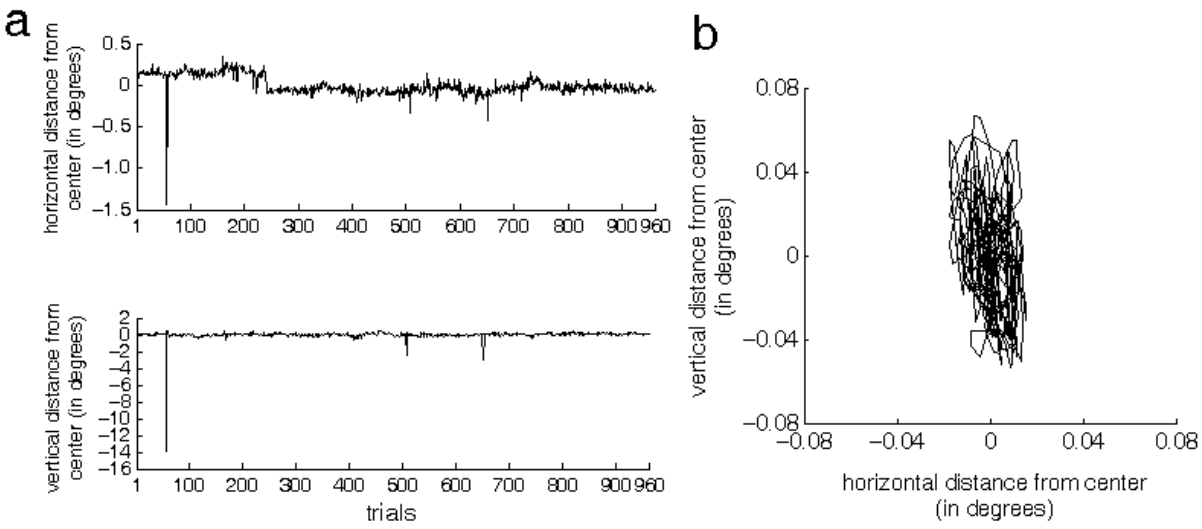

**Figure 6. Relationship between d' and criterion c in Experiment 1.6**

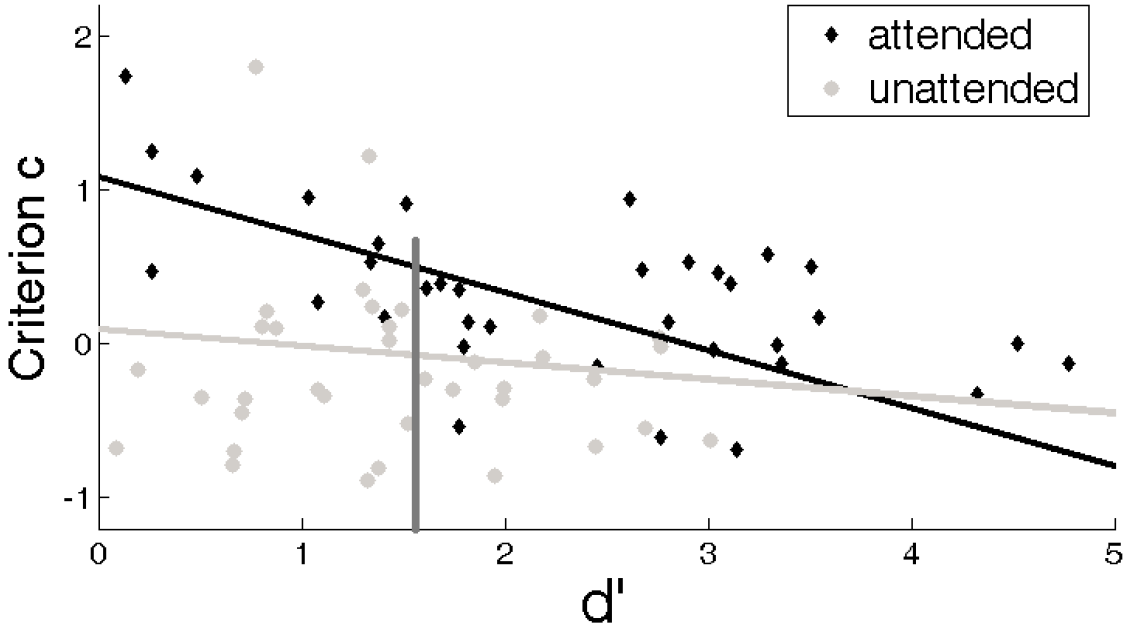

**Figure 7. Relationship between contrast level and criterion c in Experiment 1.6**

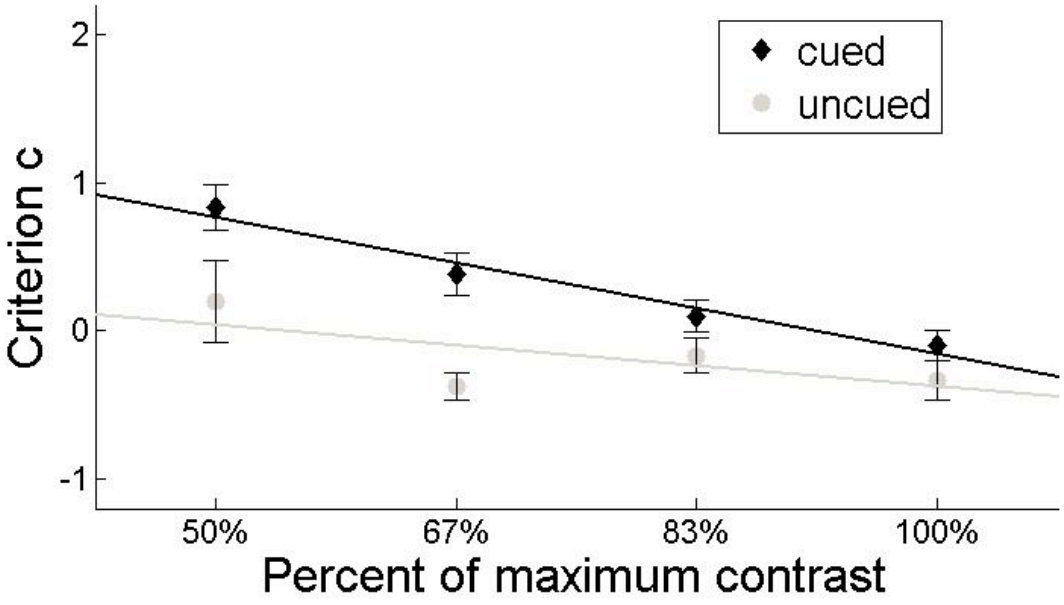

**Figure 8. Hit and false alarm rates in Experiment 1.6**

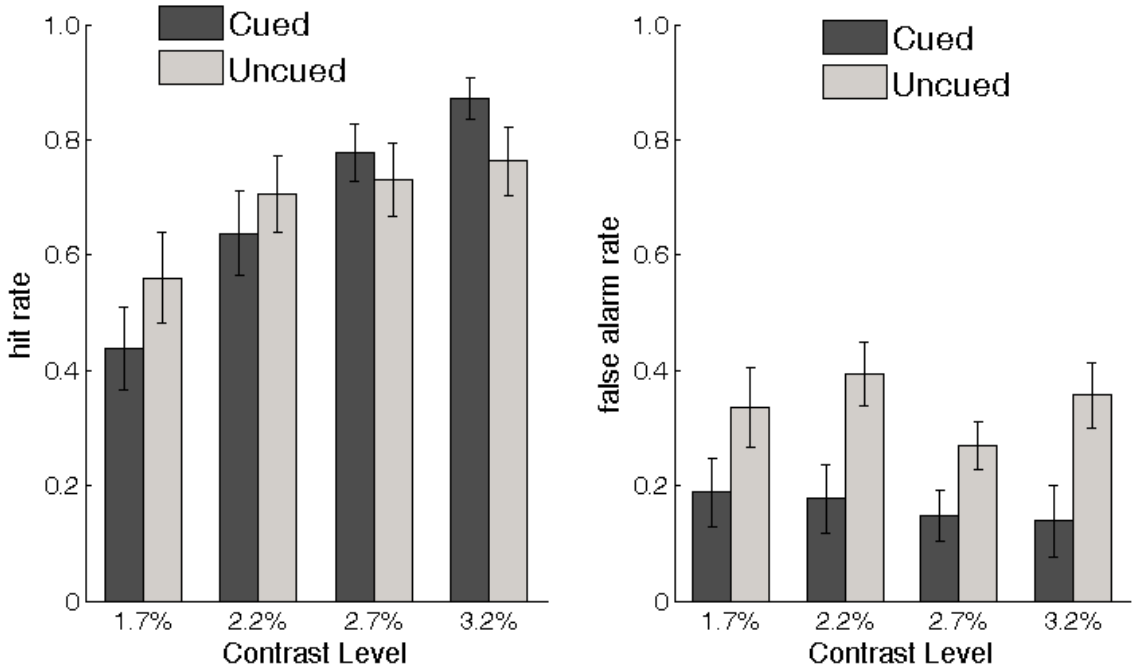

**Figure 9. Task design in Experiments 2.1 and 2.2**

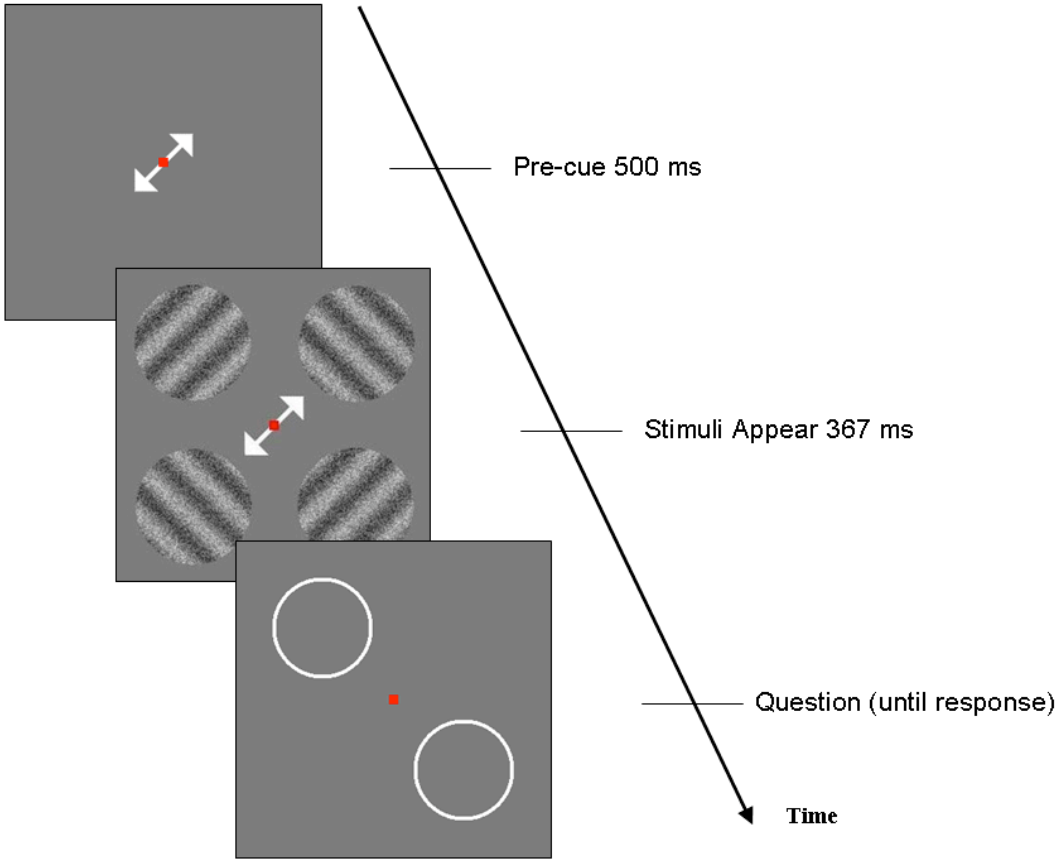

**Figure 10. Capacity d' and visibility ratings in Experiment 2.1**

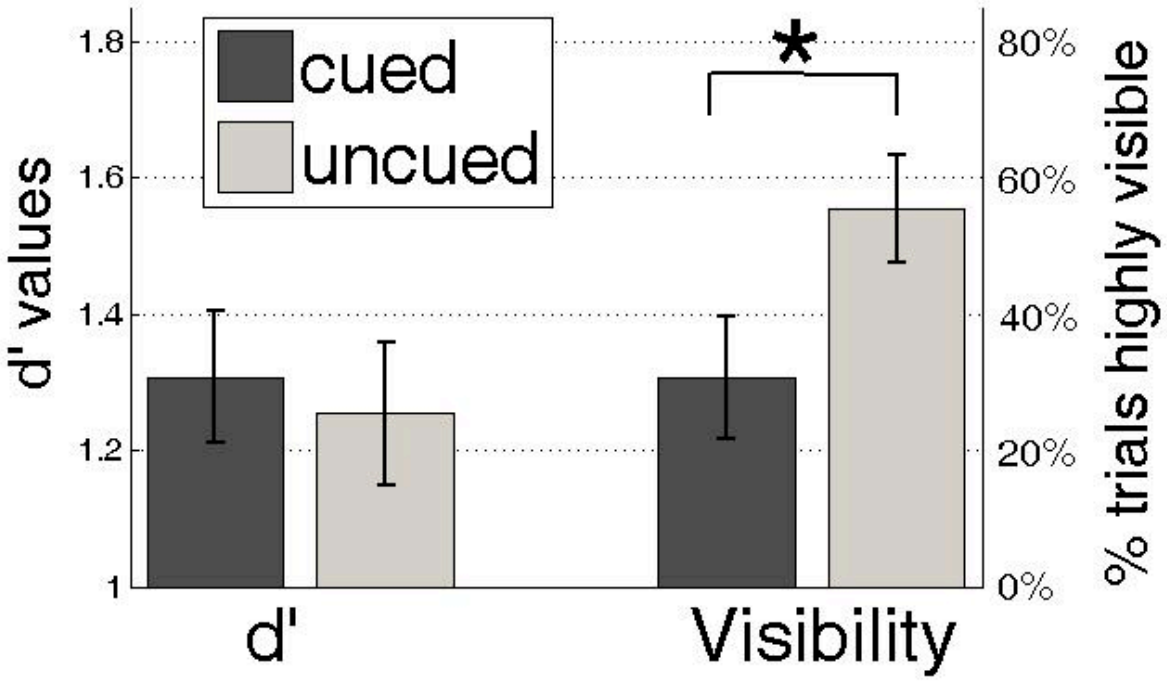

**Figure 11. Relationship between d' and visibility ratings in Experiment 2.2**

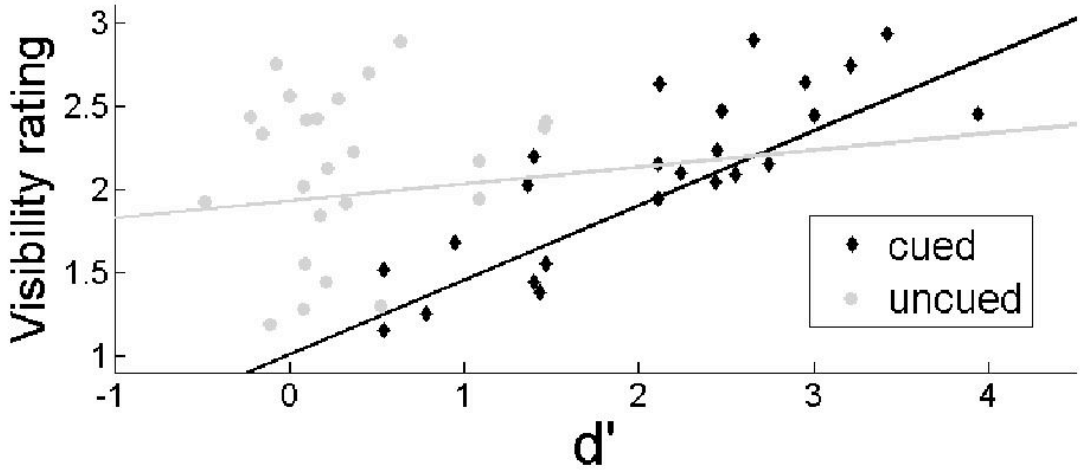

**Figure 12. Relationship between contrast and visibility ratings in Experiment 2.2**

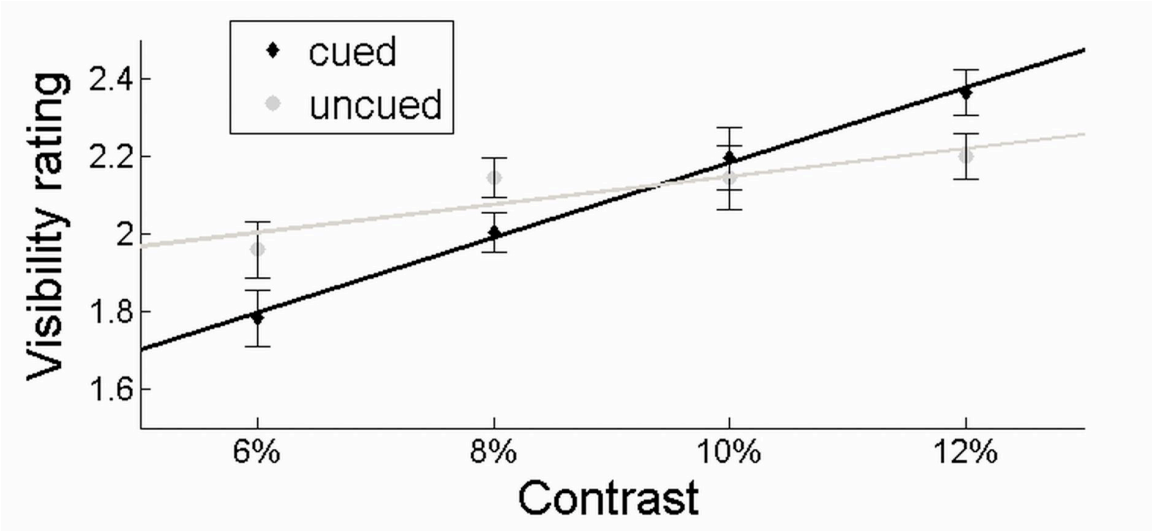

**Figure 13. Task design in Experiment 2.3**

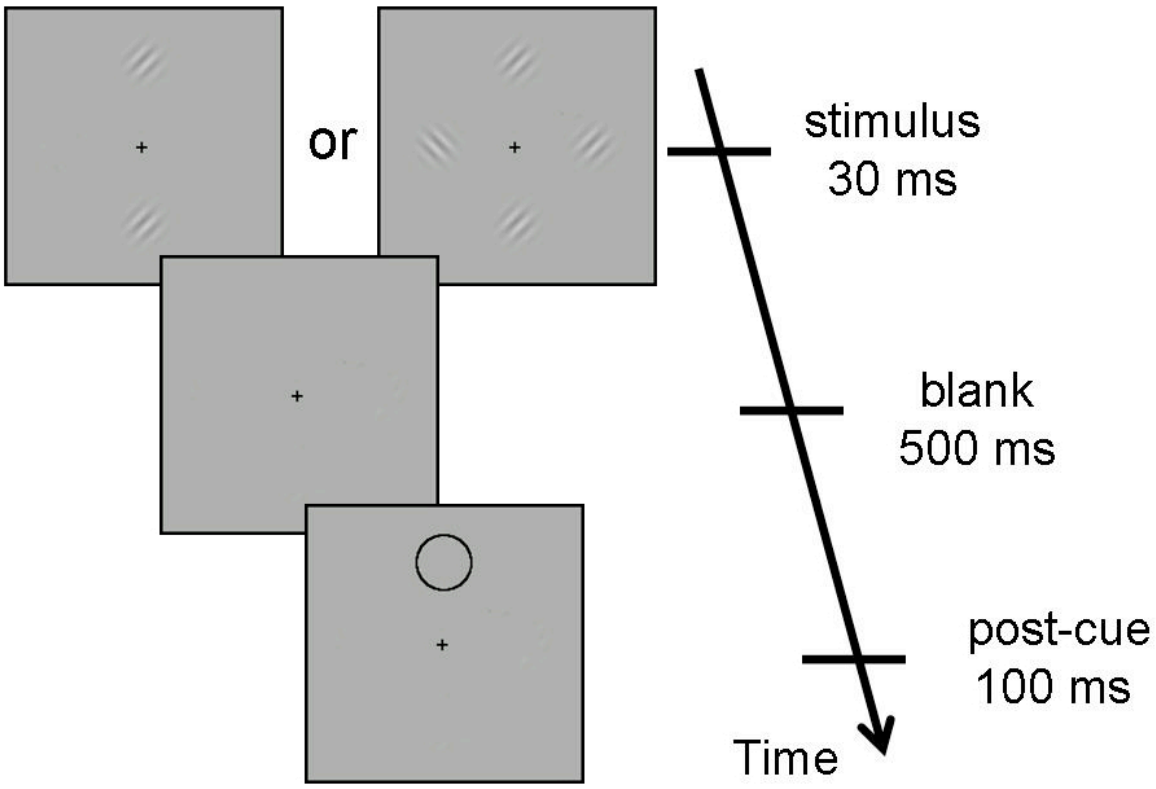

**Figure 14. Capacity d' and visibility ratings in Experiment 2.3**

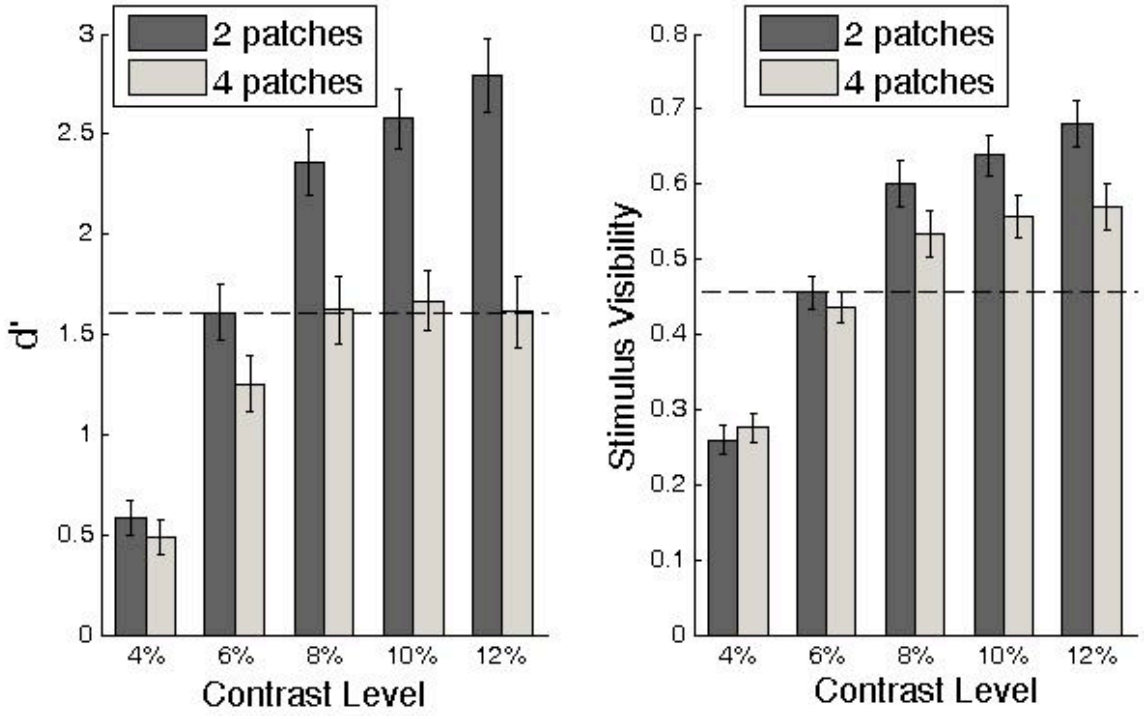

**Figure 15. A depiction of discrimination between two classes of stimuli within the framework of signal detection theory**

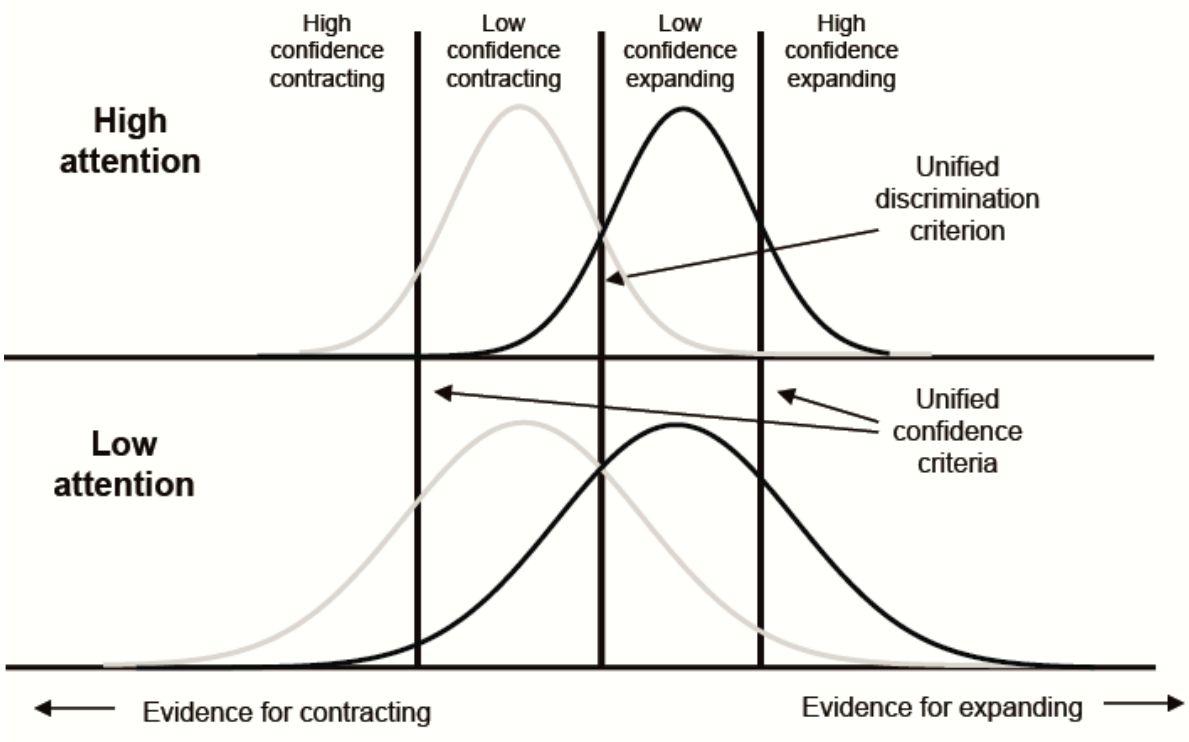

**Figure 16. Data fits of the variance reduction (VR) model**

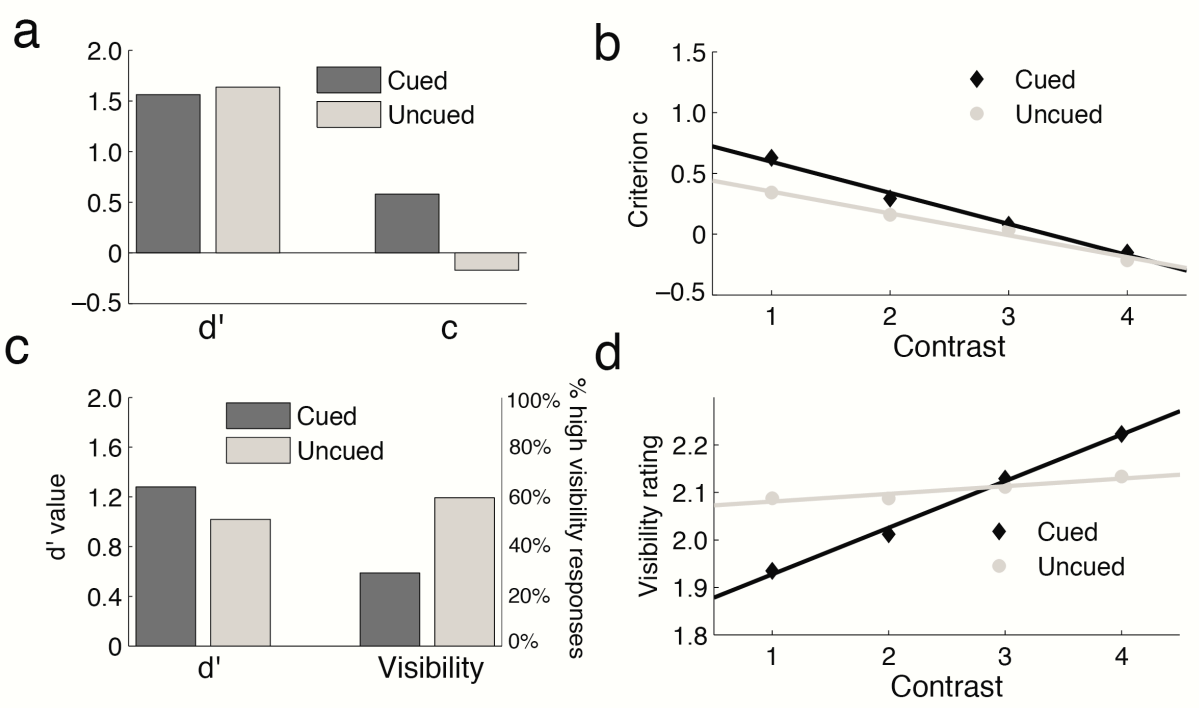

**Figure 17. A depiction of the VR model**

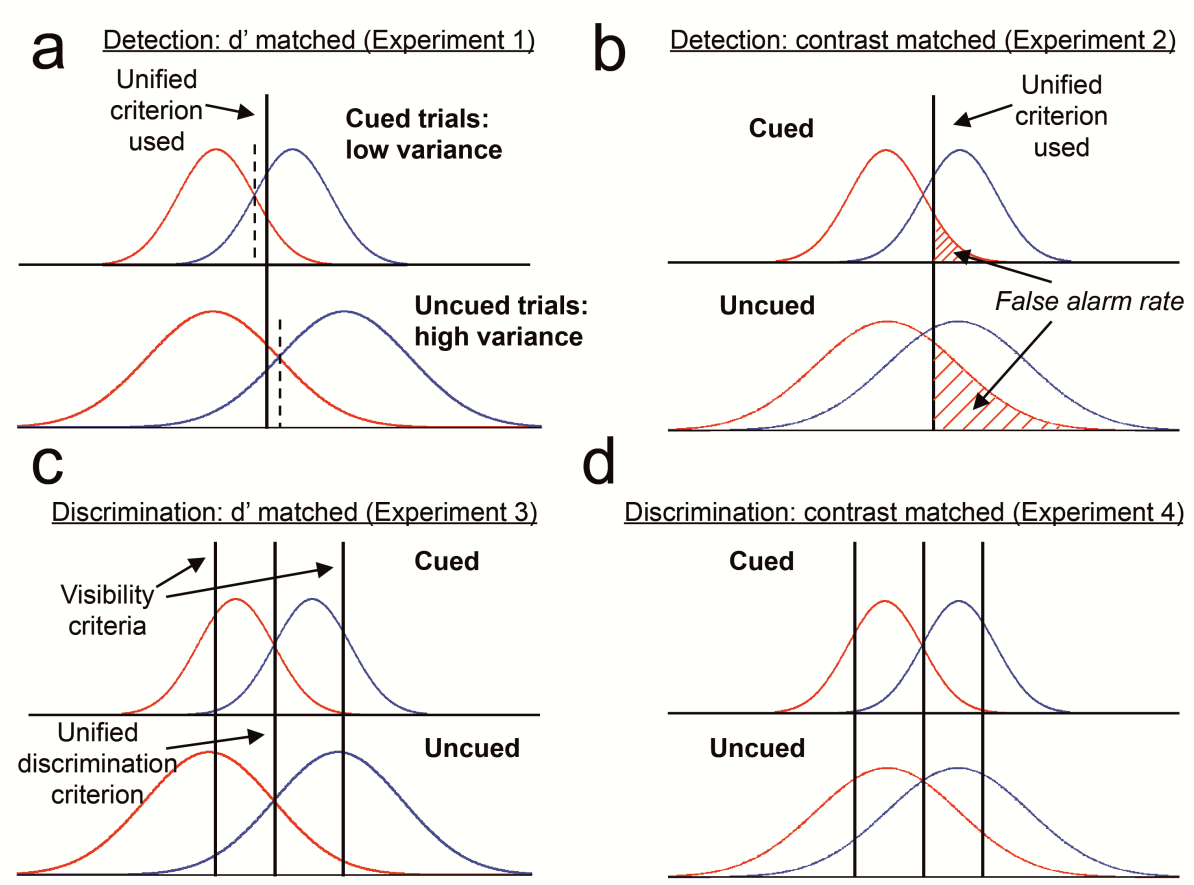

**Figure 18. Models in which attention does not affect the variability of the distributions cannot account for the observed data**

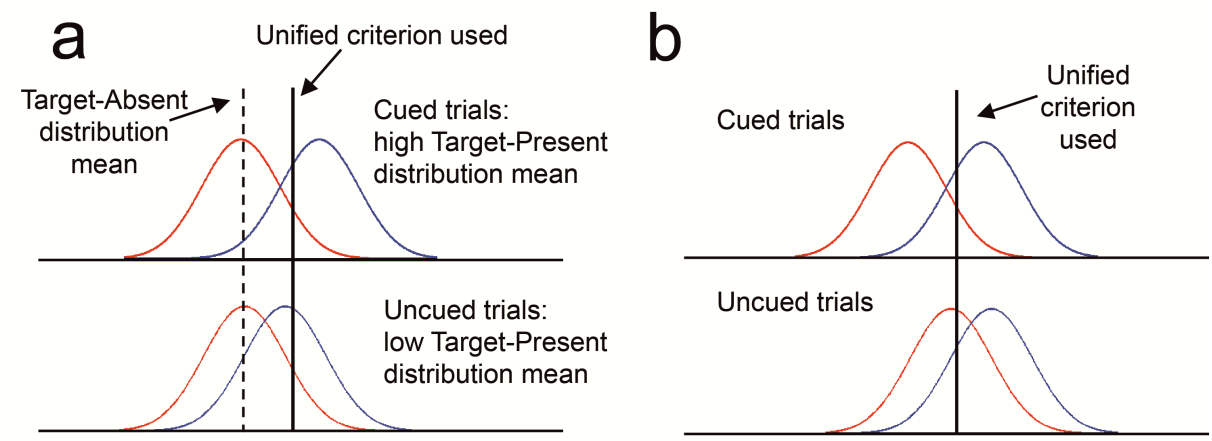

**Figure 19. Task design in fMRI experiment (Chapter 4)**

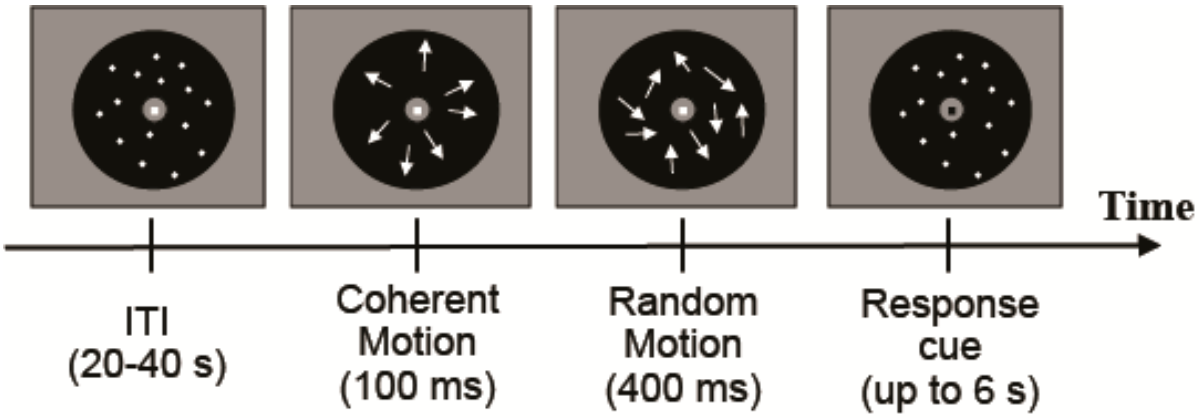

**Figure 20. Localization of the dorsal attention network (DAN)**

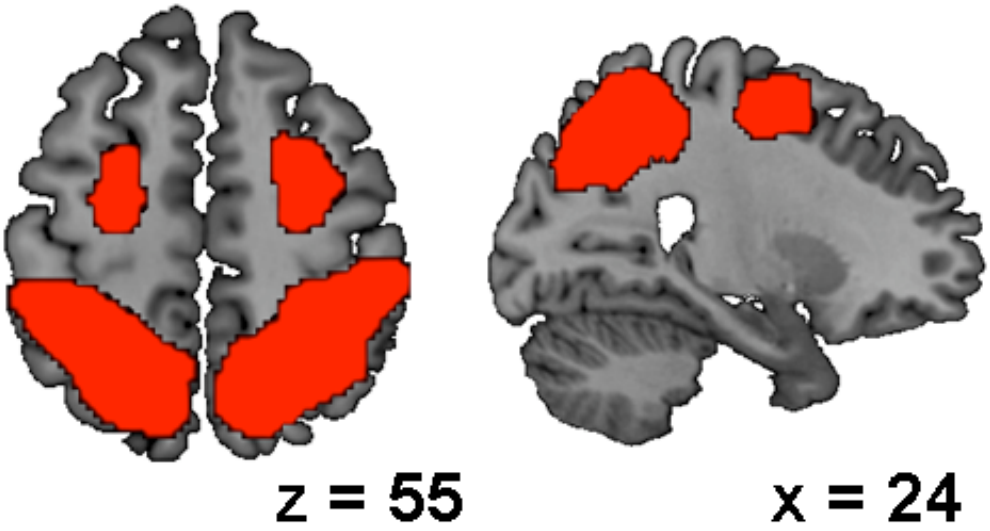

**Figure 21. Average time courses in DAN**

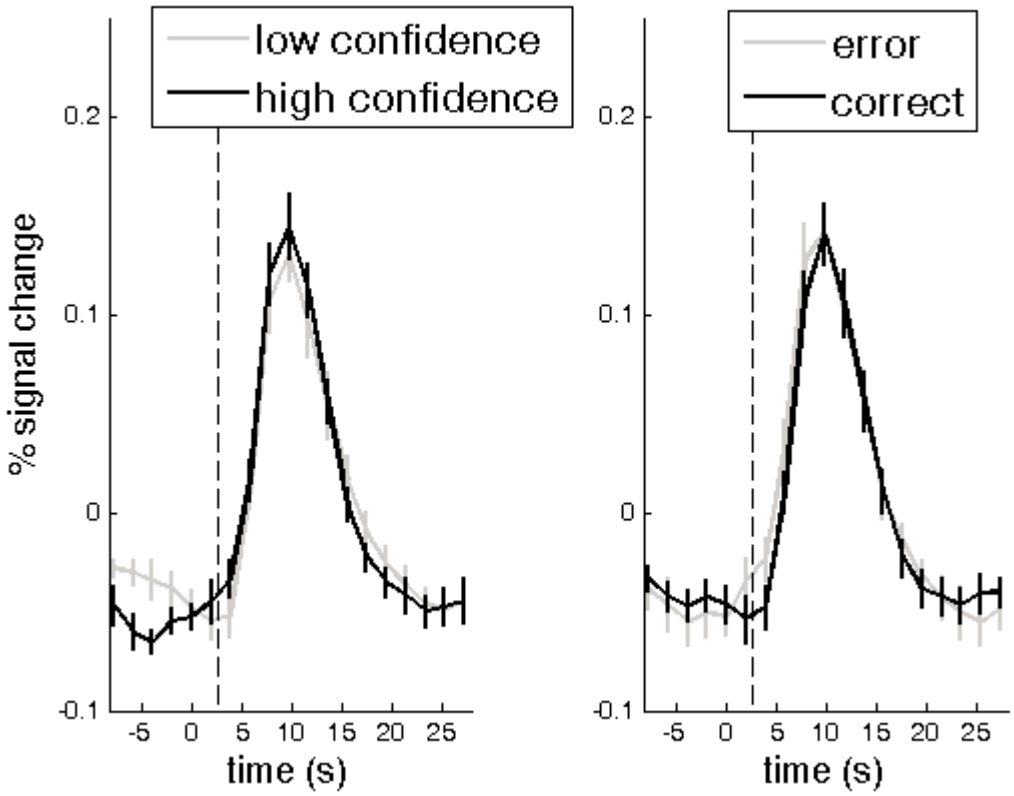

**Figure 22. Mean pre-stimulus signal for each of the four high/low attention x correct/error conditions**

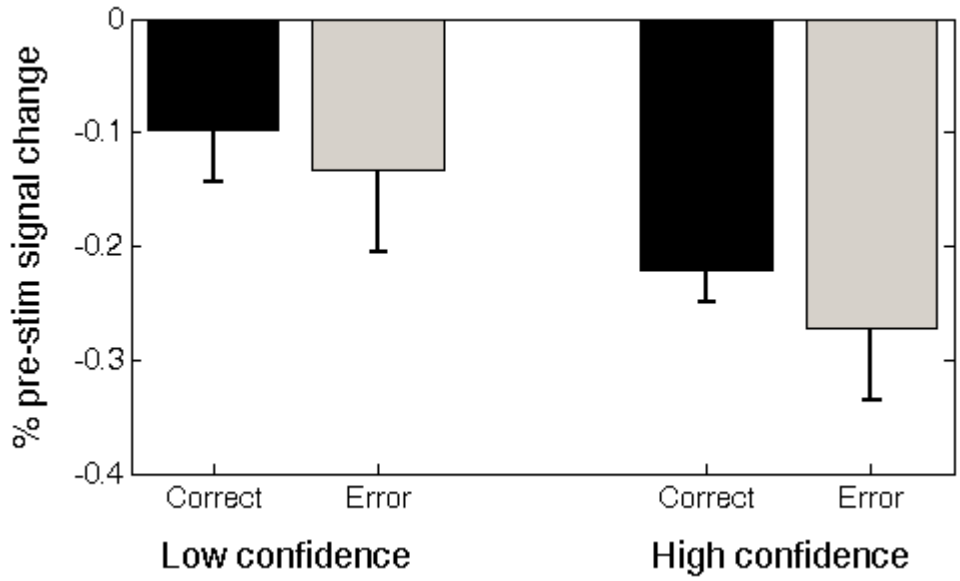

**Figure 23. Evoked activations in MT+**

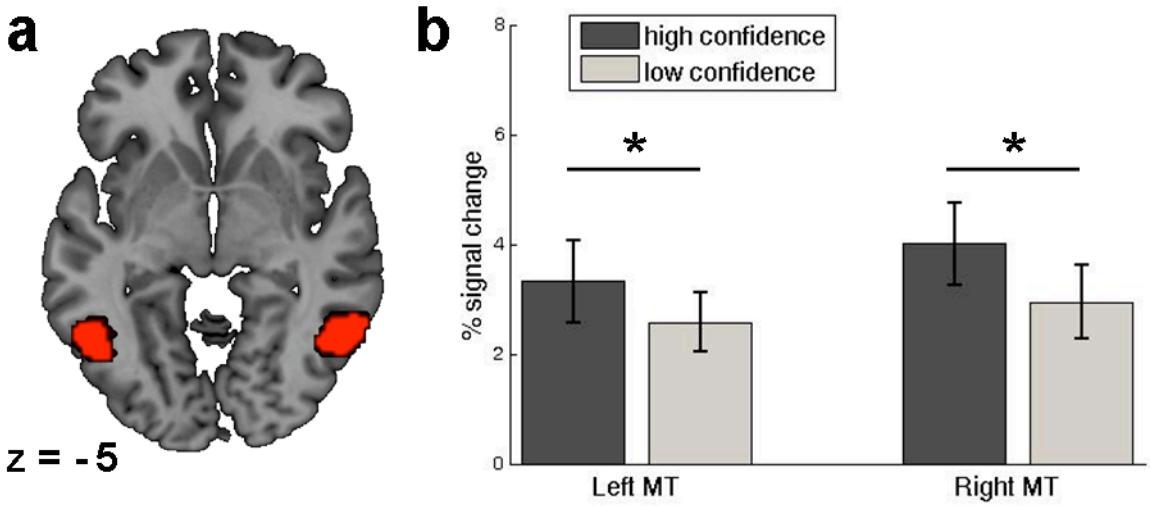

**Figure 24. Results in the default network**

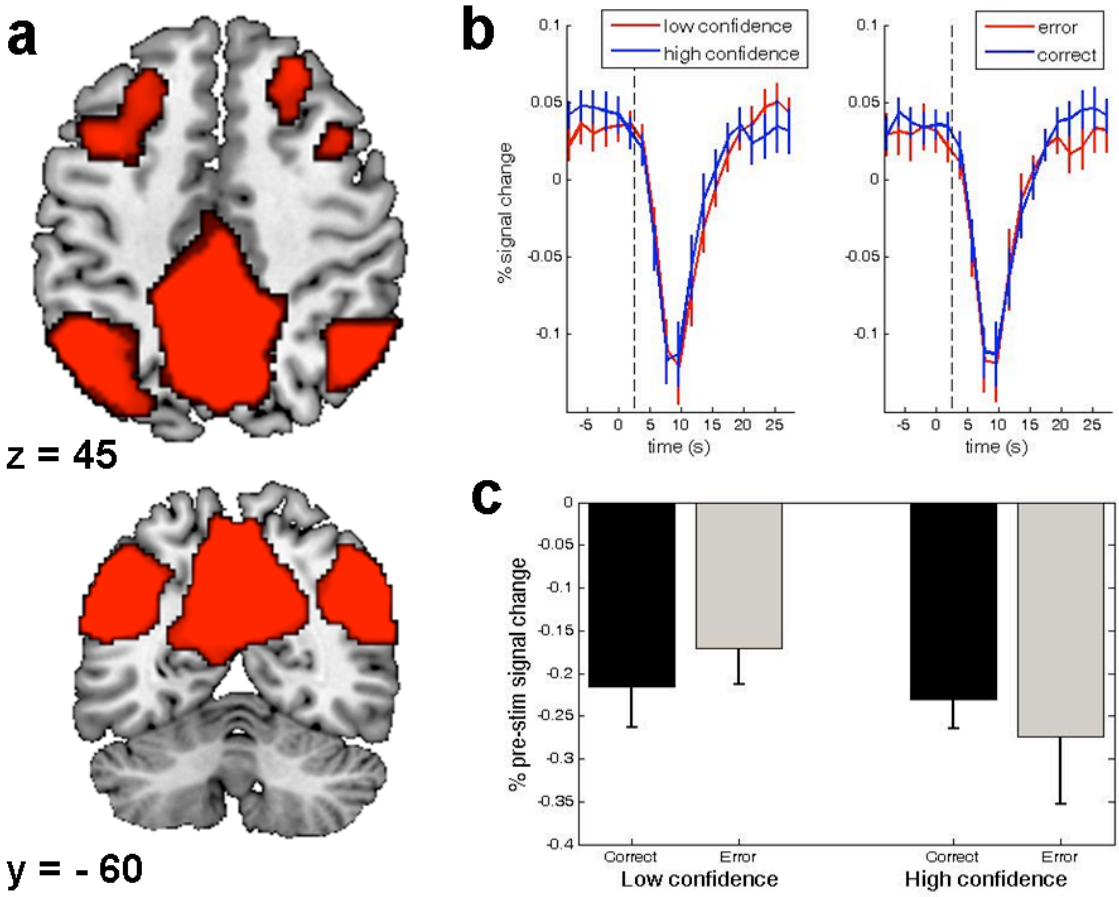

**Figure 25. Results in the alertness network**

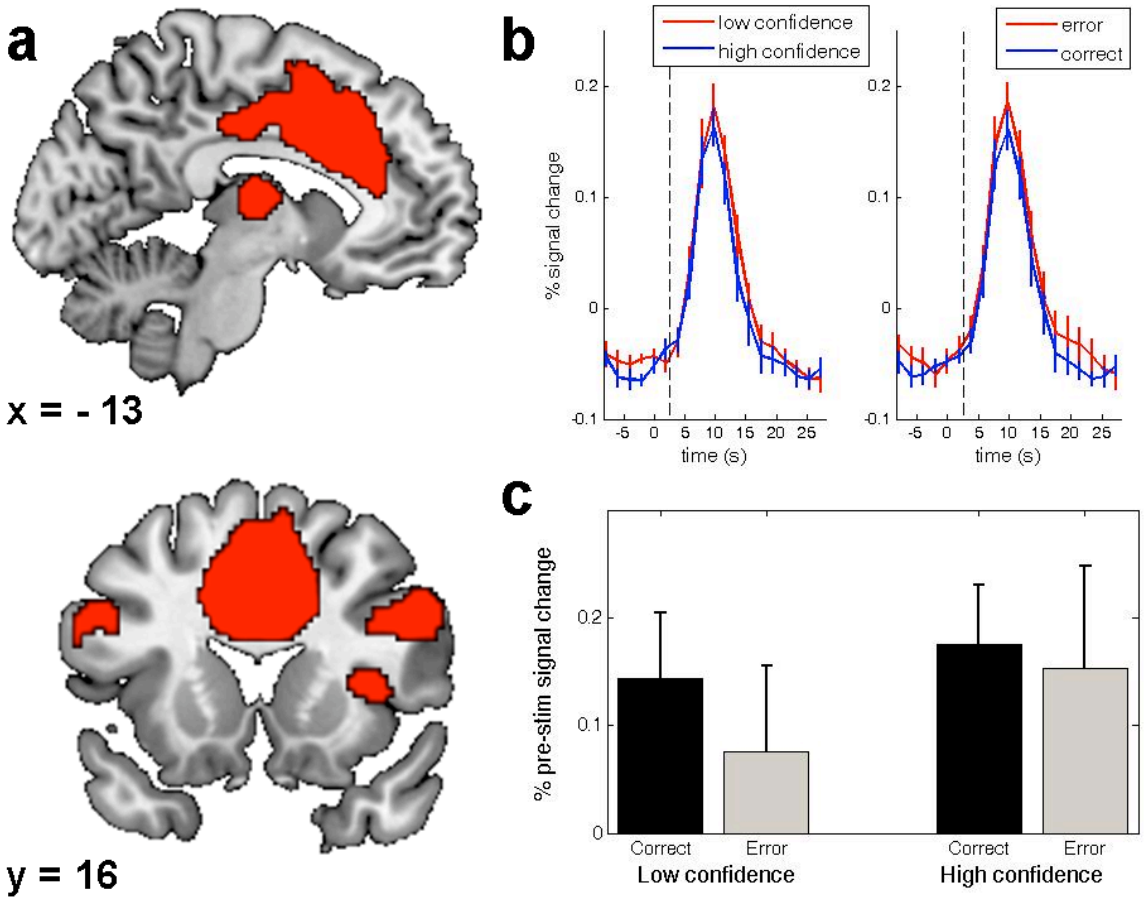

**Figure 26. Attention decreases the variability of the evoked response in MT+**

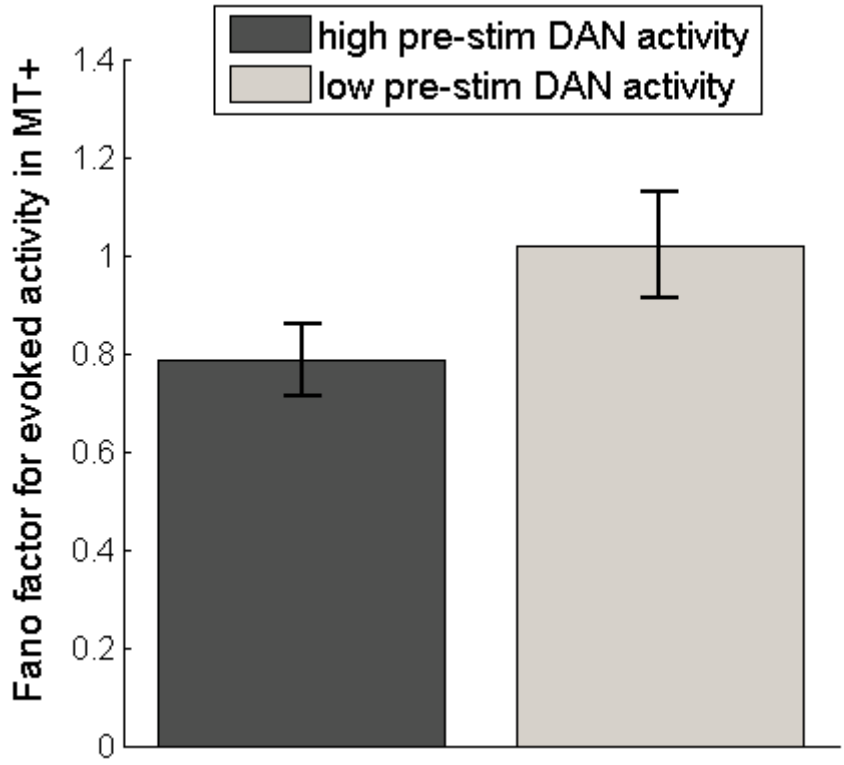

**Figure 27. Task design in the TMS experiment (Chapter 5)**

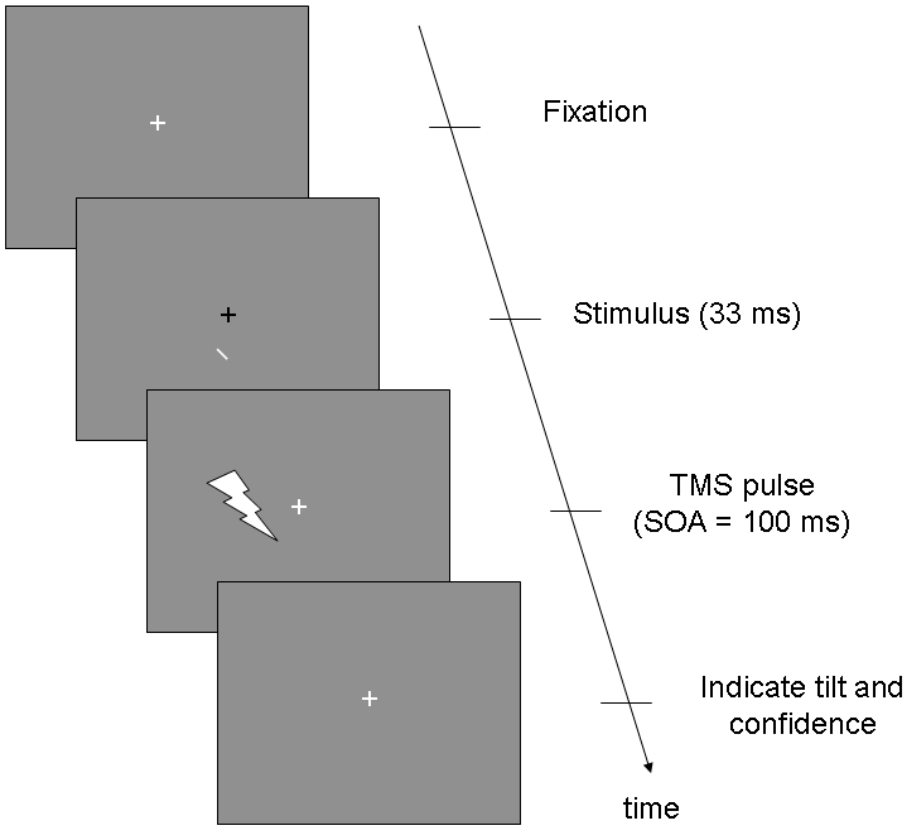

**Figure 28. Effect of occipital TMS on capacity d'**

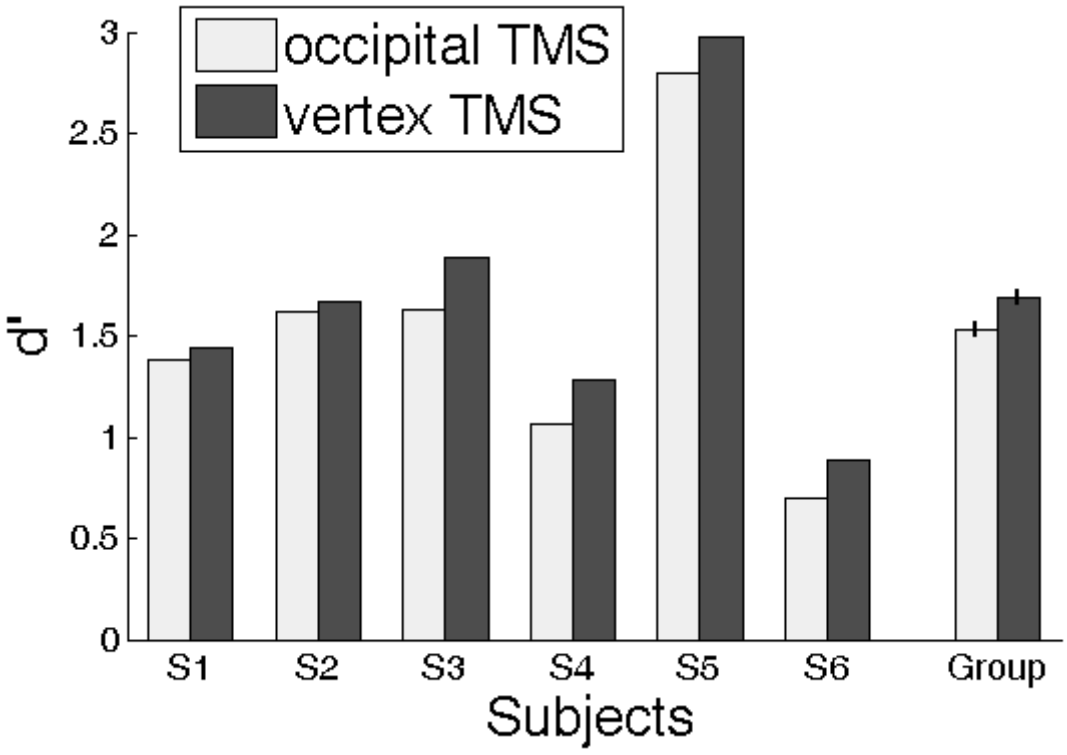

**Figure 29. Effect of occipital TMS on confidence ratings**

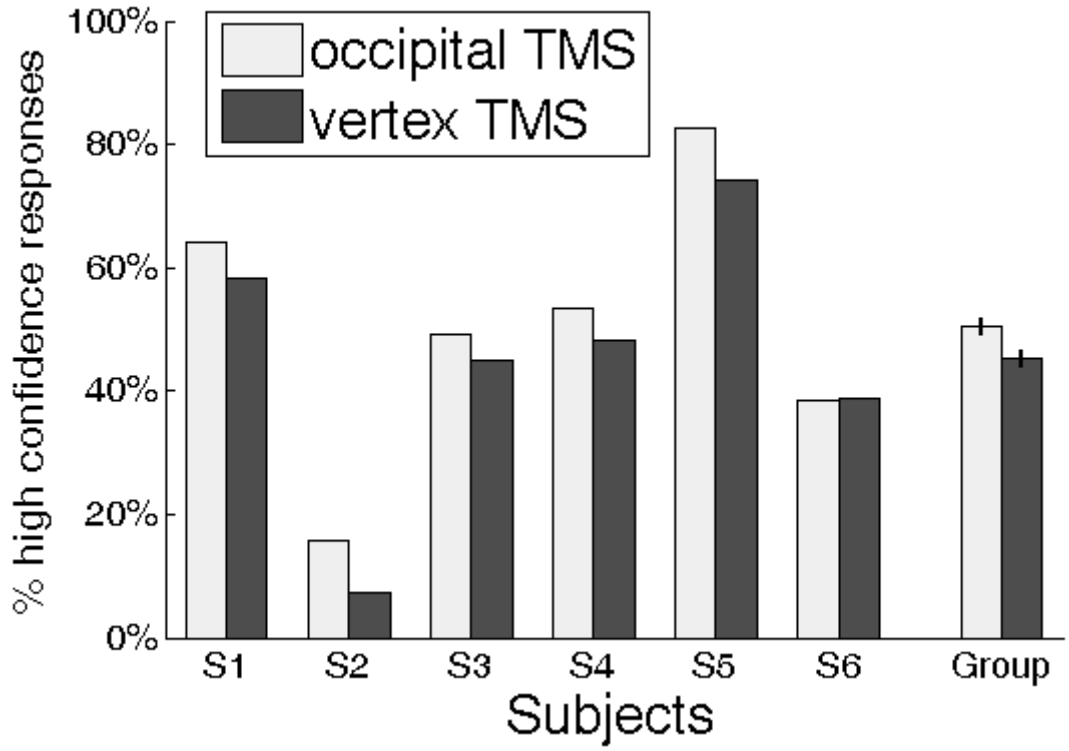

**Figure 30. A depiction of the five models used in the TMS experiment**

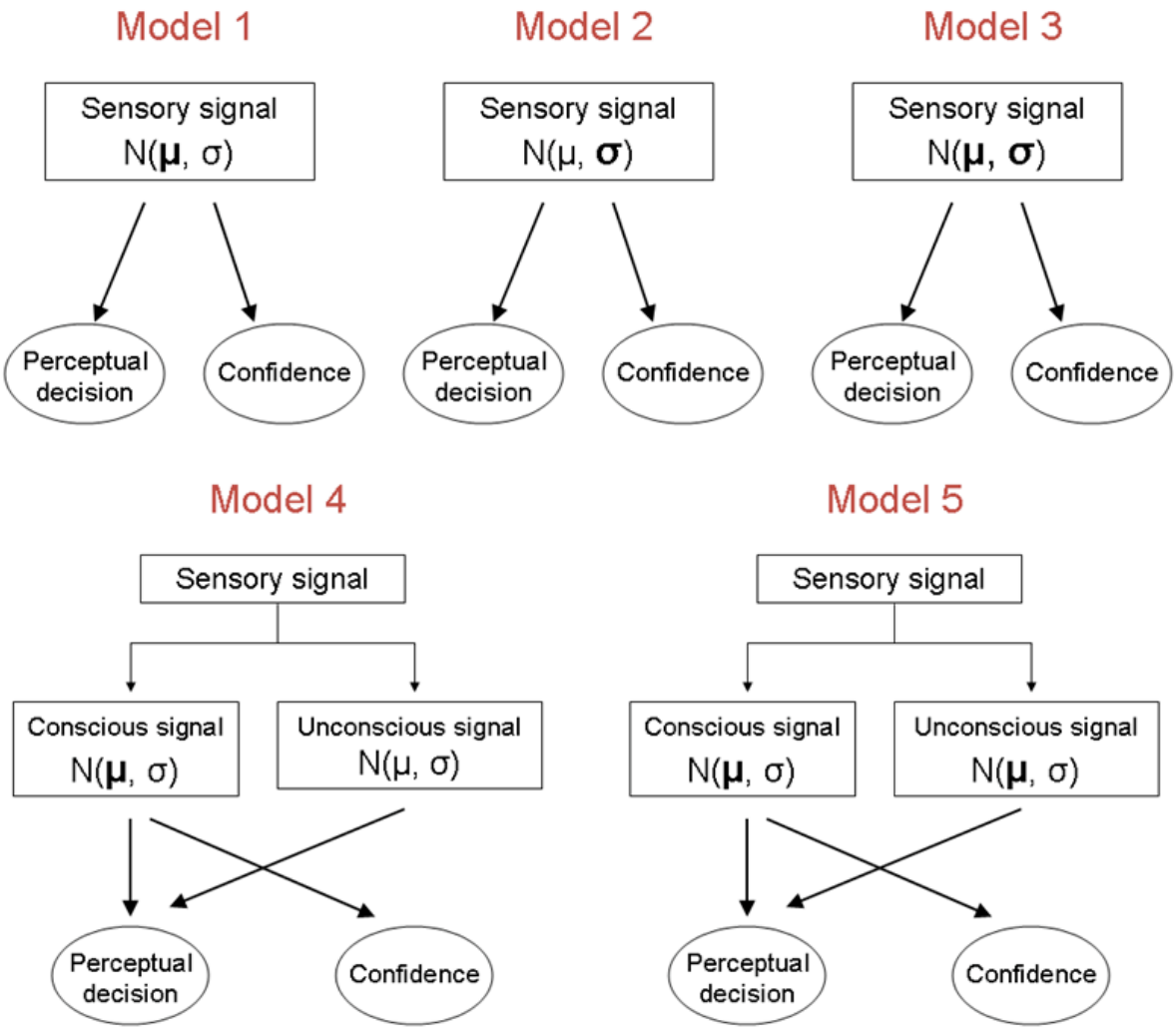

**Figure 31. Model selection results**

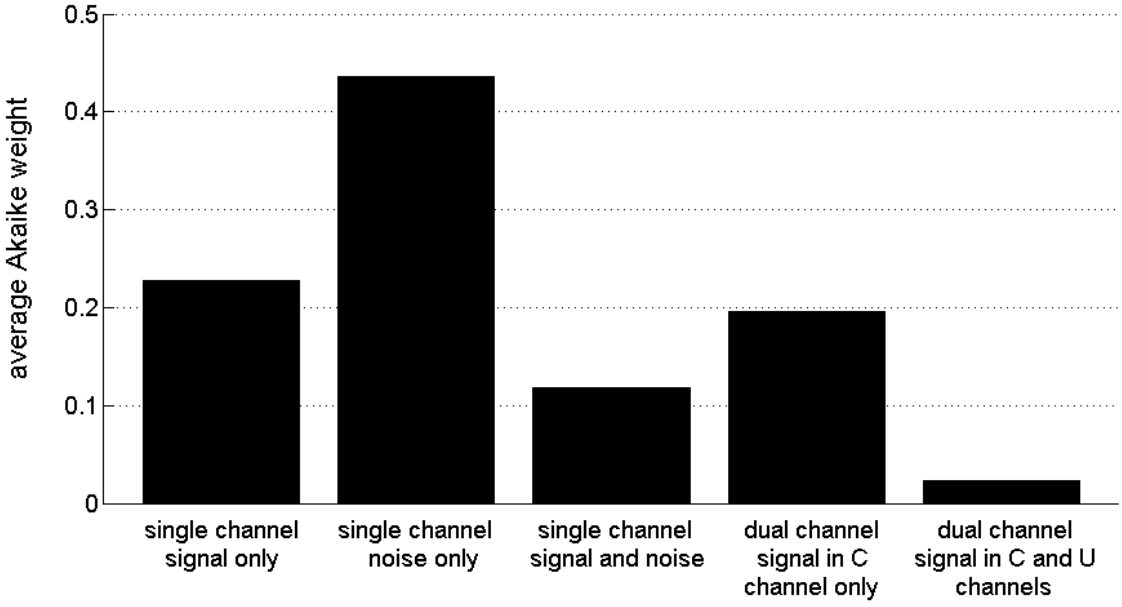

**Figure 32. Fitted data of the winning Model 2**

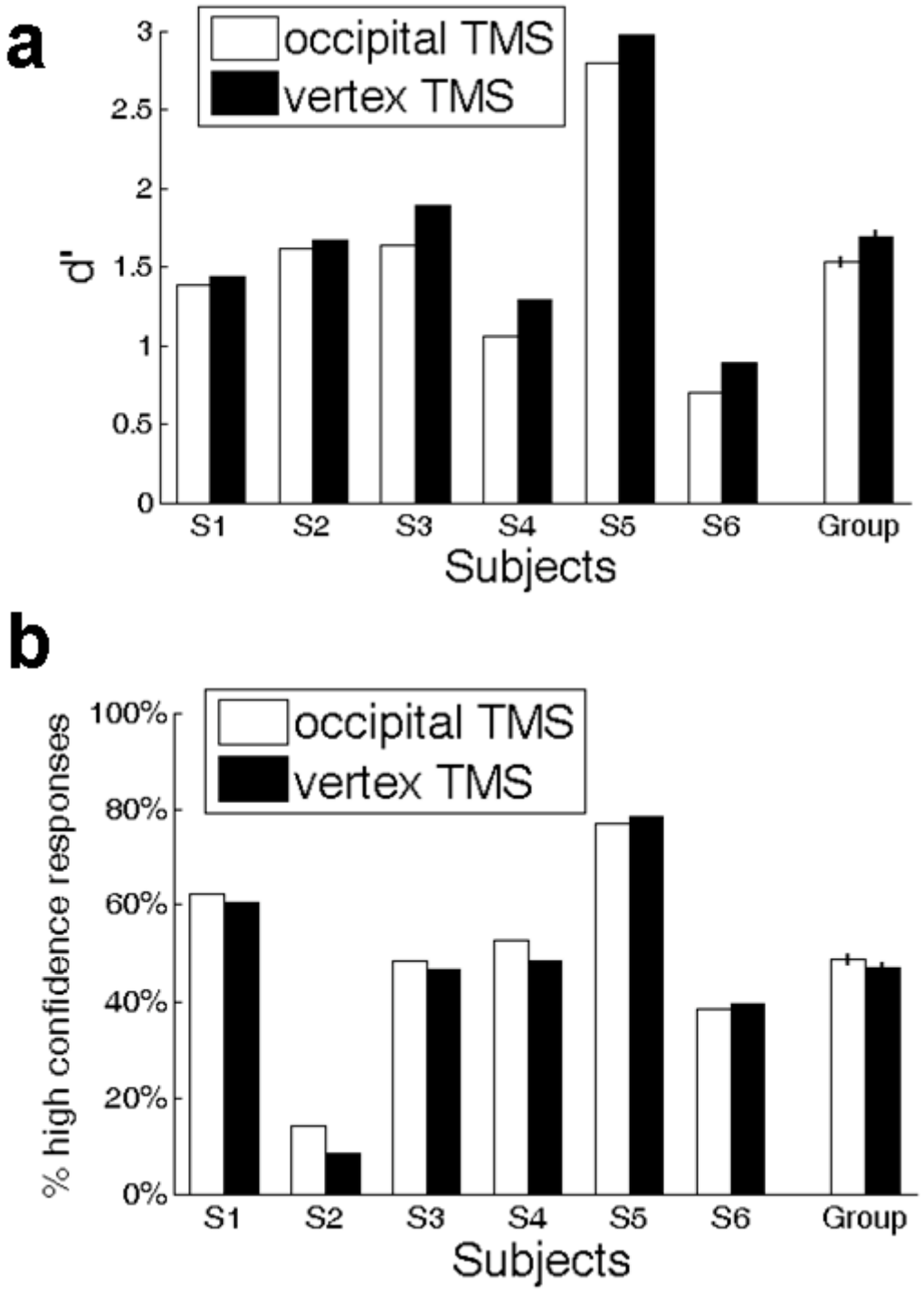

**Figure 33. Performance results in the experiment on metacognition (Chapter 6)**

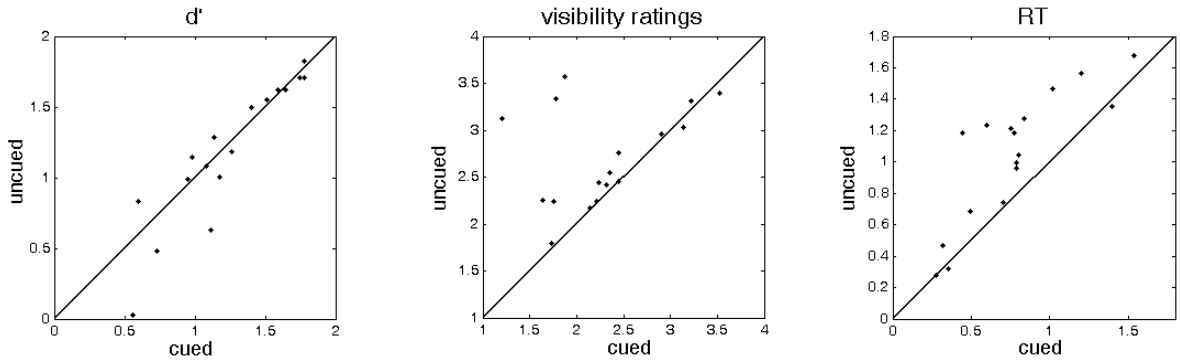

**Figure 34. Results on metacognitive sensitivity**

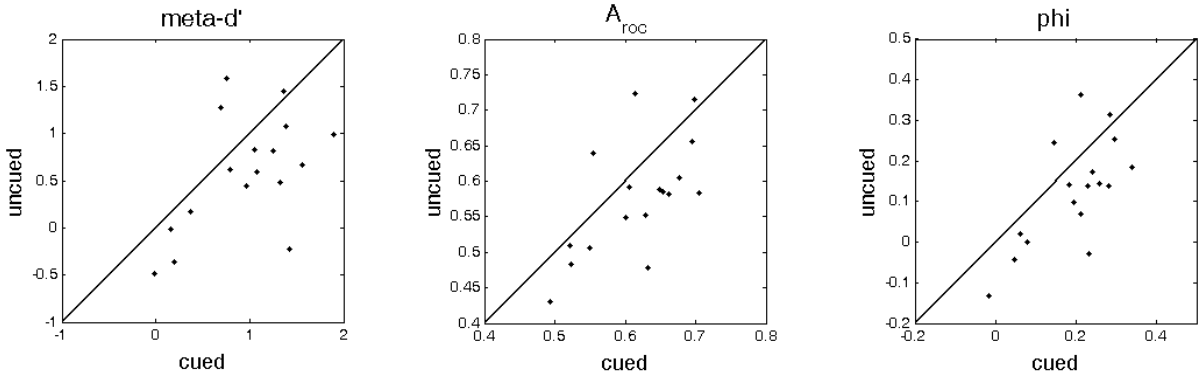

**Figure 35. Model fits of the VCJR model for performance and metacognition**

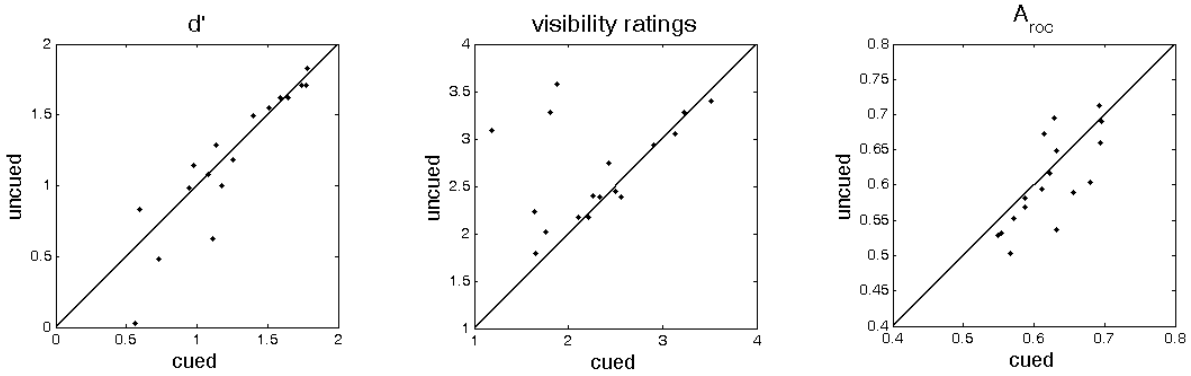

**Table 1. Pre-stimulus activity in some common brain networks**

| | % signal change | | | | Statistical effects | | |
|---|---|---|---|---|---|---|---|
| Network | High conf, correct | High conf incorrect | Low conf, correct | Low conf, incorrect | Effect on confidence | Effect on accuracy | Confidence x accuracy interaction |
| Dorsal attention network | – 0.221 | – 0.272 | – 0.098 | – 0.133 | F(1,13)=6.68 *p*=0.02 | F(1,13)=0.45 *p*=0.52 | F(1,13)=0.02 *p*=0.89 |
| Default network | 0.175 | 0.153 | 0.144 | 0.076 | F(1,13)=0.58 *p*=0.49 | F(1,13)=0.24 *p*=0.64 | F(1,13)=0.14 *p*=0.71 |
| Alertness network | – 0.23 | – 0.275 | – 0.215 | – 0.17 | F(1,13)=2.17 *p*=0.16 | F(1,13)=0 *p*=0.99 | F(1,13)=1.31 *p*=0.27 |
| Dorsal attention network (ICA) | – 0.157 | – 0.244 | – 0.052 | – 0.117 | F(1,13)=5.61 *p*=0.03 | F(1,13)=1.48 *p*=0.25 | F(1,13)=0.07 *p*=0.8 |

**Table 2. Fit from our computational simulations**

| Measure | Pre-stim DAN in low confidence trials | Pre-stim DAN in high confidence trials | Pre-stim DAN in correct trials | Pre-stim DAN in error trials | % correct trials | % high confidence trials |
|---|---|---|---|---|---|---|
| observed value | -0.129 | -0.246 | -0.172 | -0.206 | 0.626 | 0.528 |
| fitted value | -0.126 | -0.242 | -0.173 | -0.212 | 0.624 | 0.528 |

# Appendix A

Prior expectation modulates the interaction between sensory and prefrontal regions in the human brain

Dobromir Rahnev[1,2], Hakwan Lau[1,2], Floris P. de Lange[2]

[1] Columbia University, Department of Psychology, New York, USA

[2] Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behavior, Netherlands

**Abstract**

How do expectations about the identity of a forthcoming visual stimulus influence the neural mechanisms of perceptual decision making in the human brain? Previous investigations into this issue have mostly involved changing the subjects' attentional focus or the behavioral relevance of certain targets, but rarely manipulated subjects' prior expectation about the likely identity of the stimulus. Also, since perceptual decisions were often paired with specific motor responses, it has been difficult to dissociate neural activity that reflects perceptual decisions from motor preparatory activity. Here we designed a task in which we induced prior expectations about the direction of a moving dot pattern, and withheld the stimulus-response mapping until the subjects were prompted to respond. In line with current models of perceptual decision making, we found that subjects' performance was influenced by their expectation about upcoming motion direction. The integration of such information into the decision process was reflected by heightened activity in the dorsolateral prefrontal cortex. Activity in this area reflected the degree to which subjects adjusted their decisions based on the prior expectation cue. Further, there was increased effective connectivity between sensory regions (motion-sensitive area MT+) and dorsolateral prefrontal cortex when subjects had a prior expectation about the upcoming motion direction. Dynamic causal modeling suggested that stimulus expectation modulated both the feedforward and feedback connectivity between MT+ and prefrontal cortex. These results provide a mechanism of how prior expectations may affect perceptual decision making, namely by changing neural activity in, and sensory drive to, prefrontal areas.

**Introduction**

Perceptual decision making has recently received great attention by researchers (Ratcliff and McKoon, 2008; Heekeren et al., 2008; Tosoni et al., 2008; Donner et al., 2009; Ferrera et al., 2009; Egner et al., 2010; Noppeney et al., 2010). Perceptual decisions are almost always informed and heavily biased by our prior expectations. For example, we identify objects much more rapidly in a familiar context that sets up priors for object recognition (Bar, 2004; Enns and Lleras, 2008).

Most studies of perceptual decision-making have looked at the influence of expectations on perceptual inference by selectively biasing participants to process one type of information over another (Summerfield et al., 2006; Summerfield and Koechlin, 2008; Esterman and Yantis, 2009; Preuschhof et al., 2010). While this may selectively bias processing for one stimulus type, the prior probability of occurrence for a stimulus (i.e., its likelihood) is not manipulated in these studies. Rather, they manipulated the amount of attention devoted to the stimulus (i.e., its relevance, see Summerfield and Egner, 2009). One recent human imaging study that manipulated prior information (rather than relevance) in perceptual decision making (Forstmann et al., 2010) observed that expectation for left- or rightward motion selectively activated a contralateral cortico-striatal circuit that was linked to the associated behavioral response. However, as in most primate neurophysiological studies (Gold and Shadlen, 2007), in the study by Forstmann et al. perceptual priors (i.e., "expecting leftward motion") covaried with motor priors (i.e., "expecting to make a leftward response"). Therefore, it is not clear whether subjects integrated the prior information into the perceptual decision making process, or whether the cue helped them to prepare the most likely response.

In the current study, we manipulated prior probability of sensory evidence during a perceptual decision making task, while avoiding a fixed stimulus-response mapping. This allowed us to assess how perceptual prior expectations influence neural activity in sensory and prefrontal areas, and uncover the mechanisms of decision making independently from motor preparation (Bennur and Gold, 2011). Behavioral results showed that stimulus expectation changed subjects' perceptual choice, rendering subjects more biased (shifted criterion) and slightly less sensitive (decreased d-prime).

Neuroimaging results indicated that when subjects had a prior expectation about the sensory evidence, this resulted in both increased activity in dorsolateral prefrontal cortex, and increased recurrent connectivity between this area and sensory area MT+.

**Materials and Methods**

*Subjects*

Twenty-three healthy right-handed subjects took part in the experiment. All participants had normal or corrected-to-normal vision. Two subjects performed at chance in discriminating the direction of motion during the experiment and were therefore excluded from further analysis. All the analyses were done with the remaining twenty-one subjects (15 females; mean age: 21.6 years, SD = 2.6 years). A written informed consent was obtained from all subjects. The research was approved by the local ethics committee where the experiment was performed (CMO region Arnhem-Nijmegen, The Netherlands).

*Stimuli and task*

Subjects were required to indicate the overall direction of motion of white dots (density = 2.4 dots/degree$^2$; speed = 6 degrees/sec) presented inside a black annulus (outer circle radius = 10 degrees; inner circle radius = 1 degree). The motion direction was either contracting or expanding. A separate set of dots was chosen to carry the coherent motion each trial. Incoherent dots moved randomly with the same speed as coherent dots. Each dot had a lifetime between 3 and 5 frames (projector refresh rate = 60 Hz), after which it was replaced by another dot at a random location within the annulus. A small fixation square was presented for the duration of the trial and subjects were required to maintain fixation on it. The stimuli were presented on grey background (Figure 1) and were generated using Psychophysics Toolbox (Brainard, 1997) in MATLAB (MathWorks, Natick, MA).

Each trial began with the presentation of a cue that was a simple geometric shape (Figure 1). Four different shapes were used: square, diamond, triangle pointing up, and triangle pointing down. For half of the subjects the first two cues were predictive of the forthcoming stimulus while the last two were

not. For the remaining half of the subjects this relationship was inverted. To further avoid any confounds from low-level physical characteristics of the cues, each shape from either pair of shapes was counterbalanced across subjects to indicate either a contracting or expanding stimulus (this was only relevant when the shape was predictive). The predictive cue indicated the forthcoming stimulus correctly on 75% of the trials. We refer to the trials in which the subsequent stimulus was correctly predicted by the cue as "valid." In the remaining 25% of the time, the predictive cues indicated the wrong direction of motion. We refer to these trials as "invalid." The non-predictive cues were followed by expanding and contracting motion equally often. Thus the non-predictive cues are also referred to as "neutral" in the rest of the paper. Subjects were fully informed about the above contingencies and were encouraged to take the cue into account when making their perceptual decisions.

Subjects gave their response with the index fingers of their left and right hands. On each trial they were instructed which hand should be used for which response. Thus, the answer "expanding" might be mapped to the right hand button press on one trial and to a left hand button press on another trial. Importantly, the instruction only came after the offset of the stimulus thus ensuring that any accumulation of evidence during stimulus presentation was not conflated with motor preparation.

The cue was presented for 250 ms followed by a fixation dot for the same duration. Then, the motion stimulus appeared for 500 ms and finally the trial-specific response mapping appeared. Subjects were not allowed to respond before the appearance of the response mapping and had one second to respond once it appeared. Reaction times were measured from the onset of the response mapping. During the experiment, no feedback was provided. The total duration of a single trial was between 3.5 to 4.5 seconds, with the intervals sampled from a uniform distribution.

Prior to the fMRI experiment, each subject took part in a one-hour training session administered on a separate day, in which the subject practiced 882 trials. During the first half of the training session subjects were given trial-by-trial feedback, while there was no feedback during the second part of the

training in order to prepare subjects for the experiment in the scanner environment. Just before the fMRI experiment, subjects practiced for additional 5 minutes (64 trials) to remind them of the task requirements.

Based on the data from the behavioral session, three motion coherence levels were chosen for each subject. The intermediate coherence level was individually chosen to produce ~75% correct responses in the absence of a cue (mean coherence = 4.4%, SD = 0.7%). The low motion coherence was fixed at 50% of the intermediate coherence, while the high motion coherence was 150% of the intermediate coherence. The presentation of motion coherence was pseudo-randomized for each subject such that each coherence level appeared equally often with each combination of cue identity and motion direction.

The experiment consisted of two sessions of 336 trials each. Trials were grouped in 42 blocks of 8 trials. Each block consisted of trials with either predictive or non-predictive cues. The predictive and non-predictive blocks alternated, with the first block counterbalanced across subjects to be either predictive or non-predictive. Each block began with a 1 s presentation of the two geometric shapes that served as cues during the subsequent 8 trials, followed by a 1 s interval of fixation. This was done to remind subjects of the type of cue during the following block. Blocks were separated by an additional 4-8 seconds (sampled from a uniform distribution), resulting in an average block length of 40 seconds.

We recorded a localizer for the motion sensitive medial temporal (MT+) region in a separate session after the main experiment. Forty-two blocks of moving dots (block duration of 16 s) were alternated with fourteen blocks of stationary dots (block duration of 16 s), resulting in ~15 minutes scan duration.

*Behavioral analyses*

We computed the signal detection theoretic (SDT) measures d' and c (Macmillan and Creelman, 2005). The variable d' is a measure of a subject's stimulus discrimination sesnsitivity, whereas c is a measure

of a subject's bias. These measures were calculated on the basis of hit and false alarm rates. By only considering the trials in which subjects were cued that the stimulus was likely to be expanding, we computed $d'_{expand}$ and $c_{expand}$. Similarly, by only considering the trials in which the cue indicating that the stimulus was likely to be contracting, we computed $d'_{contract}$ and $c_{contract}$. Since we were not interested in general performance differences between contracting and expanding motion, we then averaged across these motion types to obtain $d'_{predictive}$, which reflects the discriminability of the motion stimuli in the context of a predictive cue. On the other hand, $c_{expand}$ and $c_{contract}$ are the biases for answering "expand" when an expanding or contracting cue was presented, respectively. Thus we expected $c_{expand}$ to be negative (corresponding to answering "expand" more than 50% of the time), and $c_{contract}$ to be positive (corresponding to answering "expand" less than 50% of the time). Criterion shift was defined as $c_{contract} - c_{expand}$, i.e., the difference of how much the cues were able to move subjects' criterion for picking one response option over the other. The bigger this shift was, the more the subject adjusted their behavior based on the predictive cue.

*fMRI acquisition*

Images were acquired on a 3 Tesla Trio MRI system (Siemens, Erlangen, Germany). Functional images were acquired using a 32-channel coil, with a single shot gradient echo-planar imaging (EPI) sequence (repetition time: 1950 ms; echo time: 30 ms; 31 ascending slices; voxel size: 3 x 3 x 3 mm; flip angle = 80 degrees; field of view = 192 mm). A high resolution anatomical image was acquired using a T1-weighted MP-RAGE sequence (repetition time: 2300 ms; echo time: 3.03 ms; voxel size: 1 x 1 x 1 mm).

*fMRI data analysis*

Analysis was performed using SPM5 (Wellcome Department of Imaging Neuroscience, London, UK). The first 6 volumes of each scan were discarded to allow for scanner equilibration. Preprocessing consisted of realignment through rigid-body registration to correct for head motion, slice timing correction to the onset of the first slice, coregistration of the functional and anatomical images, segmentation of the anatomical image, normalization to Montreal Neurological Institute (MNI) space

using the grey matter image obtained from the segmentation, interpolation of functional images to 2 x 2 x 2 mm, and smoothing with a Gaussian kernel with a full-width at half-maximum of 8 mm. A high-pass filter (cutoff 128 s) was applied to remove low-frequency signals like scanner drift.

Regressors for the first-level analysis were obtained by convolving the unit impulse time series for each condition with the canonical hemodynamic response function. We modeled the three different cue-stimulus types (valid, invalid, neutral) separately for all trials in which subjects responded correctly. Trials with incorrect or no responses were modeled as a regressor of no interest. The onsets of the trials were defined at the onset of the cue presentation, and the duration was set to one second - the period from the onset of the cue to the offset of the motion stimulus. We included 12 nuisance regressors related to head motion: three regressors related to translation and three regressors related to rotation of the head, as well as their derivatives (Lund et al., 2005).

Unless stated otherwise, the reported activations are at $p < 0.05$ corrected at the cluster level for multiple comparisons using an auxiliary (uncorrected) voxel threshold of $p < 0.001$. This auxiliary threshold defines the extent of activated clusters which formed the basis of our statistical inference (Friston et al., 1996).

*Psychophysiological interaction (PPI) analysis*

We performed psychophysiological interaction (PPI) analyses to test if connectivity between MT+ and regions in the parietal and prefrontal cortex depended on the identity of the cues. PPI represents a measure of context-dependent connectivity, explaining regionally specific responses in one brain area in terms of the interaction between responses in another brain region and a cognitive or sensory process (Friston et al., 1997). We defined the left and right MT+ for each subject using the data from the localizer by applying the contrast “moving dots > stationary dots”. We also defined the dorsolateral prefrontal cortex (DLPFC) and the intraparietal sulcus (IPS) based on the contrast “valid + invalid (i.e., predictive) > neutral (i.e., non-predictive)” in the main experiment. For each subject, BOLD response time series values were extracted from these regions. The analysis was constructed to

test for differences in the regression slope of MT+ activity on IPS and DLPFC, depending on whether subjects were engaged in tasks with either predictive or non-predictive cues.

To construct the PPI regressors we multiplied the time courses extracted from left and right MT+ ("physiological factor") with the time course of the experimental manipulation (predictive trials – non-predictive trials, "psychological factor;" Gitelman et al., 2003). This PPI regressor, alongside the task regressors and the time course of the seed region, were included in the statistical model for each subject. Separate models were created for left and right MT+. To visualize the regression slopes separately for predictive and non-predictive trials, we multiplied the time course of neural activity in MT+ with task regressors relating to predictive and non-predictive trials, respectively. However, it should be noted that the difference in slope values between conditions, rather than their absolute values, are of interest. The reason is that the slope for predictive trials is implicitly biased by the non-predictive trials and vice versa.

Average parameter estimates for each region of interest (ROI) were calculated for each subject separately. These estimates were then subjected to a paired-samples t-test to test the prediction that connectivity between MT+ and left DLPFC and/or IPS is different for predictive compared to non-predictive blocks.

*Dynamic Causal Modeling (DCM) analysis*

Dynamic causal modeling aims to model directed (causal) influences between regions (Friston et al., 2003). In the PPI analysis, we observed stronger connectivity between MT+ and the left dorsolateral prefrontal cortex (DLPFC). DCM allows us to adjudicate between stronger feedforward, feedback or recurrent interactions between MT+ and DLPFC. This analysis was performed using SPM8 (Wellcome Department of Imaging Neuroscience, London, UK). To perform the DCM analysis we estimated a first level analysis with the following regressors: stimulus (modeling the main effect of stimulus, i.e., stimulus > baseline) and expectation (modeling the main effect of expectation, i.e., predictive trials > non-predictive trials). To extract the time courses for each of the regions of interest,

we thresholded the subject-specific statistical F maps for the effects of no interest from the first level model described above at $p < 0.01$ and masked the resulting activations with the subject-specific MT+ regions obtained from the localizer or with the left DLPFC region obtained from the group analysis above.

The main goal of the DCM analysis was to investigate whether the modulation of the connection between MT+ and left DLPFC by expectation was primarily feedforward, feedback, or recurrent. Thus, we specified three different models that differed in whether expectation influenced feedforward, feedback or both connections between MT+ and DLPFC. All models included intrinsic bilateral connections from left and right MT+ to left DLPFC (Figure 6A). Furthermore, we considered three different possibilities for how the stimulus itself influenced activity in and connectivity between the nodes of the network. We constructed models in which the stimulus only modulated MT+, models in which the stimulus modulated both MT+ and DLPFC, and models in which the stimulus modulated MT+, DLPFC, and the connectivity between MT+ and DLPFC. Together, this resulted in 9 models (feedforward, feedback, and recurrent models, each with three possible implementations of how stimulus affected activity). We fitted each of these nine models for each subject separately. Then, using a hierarchical Bayesian approach (Stephan et al., 2009) we compared the feedforward, feedback, and recurrent families of models by computing the exceedance probability of each class of models. The exceedance probability in this analysis reflects the likelihood that a class of models explains the data better than the other class of models in the comparison. We also used Bayesian model comparison to compare all individual models across subjects at the second level.

**Results**

*Behavioral results*

The identity of the cues (valid, invalid, or neutral) significantly influenced the proportion of correct responses ($F(2,40) = 12.63$, $p < 0.001$). In particular, the proportion of correct responses was higher when the cue was valid rather than neutral ($t(20) = 3.09$, $p = 0.006$, average difference = 5%), and lower when the cue was invalid rather than neutral ($t(20) = -3.30$, $p = 0.004$, average difference =

14%). Cue identity also significantly influenced reaction times (RTs; $F(2,40) = 12.57$, $p < 0.001$), though the pattern was more complex (Figure 2A). The level of motion coherence influenced both the proportion of correct responses ($F(2,40) = 95.94$, $p < 0.001$) and RTs ($F(2,40) = 10.24$, $p < 0.001$) with higher motion coherence leading to faster and more accurate responses. There was also an interaction between motion coherence and cue identity for both accuracy ($F(4,80) = 3.93$, $p = .013$) and RT ($F(4,80) = 5.23$, $p = .005$).

We investigated the effect of the predictive cues on discrimination sensitivity, as measured with the signal detection measure d' (Macmillan and Creelman, 2005). The idea behind this analysis was to estimate subjects' capacity to perform the motion discrimination task, either with or without an expectation about upcoming motion direction. We computed $d'_{predictive}$ using all the valid and invalid trials and $d'_{neutral}$ using all the neutral trials. The key point to note about the procedure is that to compute $d'_{predictive}$ we separated the predictive (valid and invalid) trials into two categories: one in which the cue indicated that the forthcoming stimulus was likely to be expanding ($d'_{expand}$) and one in which the cue indicated a likely contracting stimulus ($d'_{contract}$). Then, $d'_{predictive}$ was simply the average of $d'_{expand}$ and $d'_{contract}$ (see Method for more details). Similarly, we computed the signal detection measure of bias c for predictive ($c_{predictive} = c_{contract} - c_{expand}$) and neutral ($c_{neutral}$) trials.

Figure 2B shows that, not surprisingly, d' increased with motion coherence ($F(2,40) = 88.63$, $p < 0.001$). More interestingly, the cues also influenced d' ($F(1,20) = 8.66$, $p = 0.008$) such that predictive (valid and invalid) cues *decreased* d'. This suggests that although subjects benefited from the predicitive cue to improve overall accuracy, the underlying stimulus processing sensitivity (independent of the aid provided by the cue) became worse after the presentation of a predictive cue. The decrease in d' for predictive cues became bigger with increasing motion coherence with the medium and high motion coherence showing a significant d' difference (medium coherence: d' difference = .36; $t(20) = 2.69$, $p = 0.014$; high coherence: d' difference = .47; $t(20) = 2.59$, $p = 0.018$) unlike the lowest motion coherence which did not show a significant d' decrease (d' difference = .13; $p = .23$).

Finally, the predictive cues significantly shifted the criterion (c: $F(1,20) = 11.89$, $p = 0.003$) demonstrating that subjects did incorporate the information of the cues. On the other hand, motion coherence had no effect on c ($F(2,40) = 1.21$, $p = 0.31$).

*Neural activity*

We first identified regions whose activity during the perceptual decision was modulated by expectation. More specifically, we investigated whether there was larger activity during trials with predictive cues (valid and invalid) compared to trials with non-predictive cues (neutral). This comparison revealed larger activity for predictive trials bilaterally in the intraparietal sulcus (IPS) and dorsolateral prefrontal cortex (DLPFC). The anterior part of the left IPS region extended into the gyral surface posterior to the postcentral sulcus, while its most posterior part reached just anterior and dorsal of the transverse occipital sulcus. The right IPS activation was smaller and did not extend as laterally and anteriorally as the activation in left IPS. In view of their anatomical locations, these activations are most properly refered to as "anterior IPS" (Shulman et al., 1999). The anatomical location of DLPFC falls close to cyto-architectonically defined Brodmann Area 46 (Rajkowska and Goldman-Rakic, 1995). These regions are shown in Figure 3A and details are provided in Table 1. We next tested for differences in these regions as a function of cue type (valid vs invalid), as well as belief updating (as indexed by criterion shift induced by the predictive cue).

Figure 3B shows the % signal change associated with the three cue types (invalid, valid, neutral), for each of the four regions. Interestingly, there was larger activity during invalidly cued trials than validly cued trials in the left ($t(20) = 2.49$, $p = 0.01$) and right DLPFC ($t(20) = 2.54$, $p = 0.01$). This activity difference was not present in either left or right IPS ($p > 0.8$). In Figure 3C we have plotted the time courses for each trial type in each of the regions of interest.

Next, we tested if inter-individual differences in the updating of one's belief on the basis of the predictive cue during the perceptual decision was related to activity levels in the expectation-related

regions of interests (IPS, DLPFC), as well in the sensory region MT+. Specifically, we investigated whether between-subject variations in brain activity differences between predictive and non-predictive trials were correlated with the extent to which subjects incorporated the predictive cues (criterion shift). Criterion shift was formalized by each individual subject's propensity to adjust his or her decision criterion based on the predictive cues (i.e., $c_{contract}$ - $c_{expand}$, see Methods). We computed correlation coefficients between these measures for bilateral DLPFC, IPS and MT+ (Figure 4). Since three subjects used the cue particularly extensively and could potentially drive some of the correlations, we used a non-parametric correlation test (Spearman's rank correlation), which is unaffected by outliers. There was a positive correlation between criterion shift and brain activity in left DLPFC (Spearman's rho = .50, $p = 0.01$) and left IPS (Spearman's rho = .61, $p = 0.002$). In other words, subjects who used the cues more extensively also showed higher activity in parietal and prefrontal regions in the left hemisphere. Conversely, criterion shift showed a negative correlation with right MT+ (Spearman's rho = -.38, $p = 0.04$).

*Effective connectivity*

The previous analysis suggests that expectation modulated neural activity in dorsolateral prefrontal cortex as a function of the bias introduced by expectation. We next tested whether the interaction between sensory (MT+) and dorsolateral prefrontal cortex was also affected by expectation. In particular, we investigated whether activity changes in MT+ was associated with larger activity changes in IPS and DLPFC during trials in which subjects had a prior expectation about the forthcoming stimulus than during trials in which subjects did not have a prior expectation about the stimulus. Indeed we observed that when subjects had an expectation, there was a stronger effective connectivity between MT+ and left DLPFC (left MT+: $t(20) = 2.19$, $p = 0.02$; right MT+: $t(20) = 2.79$, $p = 0.006$) than when subjects had no expectation about the upcoming motion direction (Figure 5). These effects were observed neither in the right DLPFC ($p > 0.07$) nor in the left or right IPS ($p > 0.14$).

The effective connectivity uncovered by the above analysis could be the result of connectivity that is primarily feedforward, primarily feedback, or recurrent (both feedforward and feedback). Therefore, we performed dynamic causal modeling (DCM) analysis in order to differentiate between these possibilities. We constructed families of models in which expectation modulated the feedforward, the feedback, or both the feedforward and feedback connections between MT+ and DLPFC (see Figure 6A). Each family consisted of individual models in which the stimulus was allowed to modulate (1) only MT+, (2) MT+ and DLPFC, and (3) MT+, DLPFC, and the recurrent connections between MT+ and DLPFC (see Methods). This produced a total of nine models.

Using a hierarchical Bayesian approach (Stephan et al., 2009) we compared the feedforward, feedback, and recurrent families of models by computing the exceedance probability of each class of models. The exceedance probability in this analysis reflects the likelihood that a class of models explains the data better than the other class of models in the comparison. The results showed that the recurrent family of models explained the data best (Figure 6B). This family had exceedance probability of 80%. The feedforward-only family of models had exceedance probability of 18%, while the feedback-only class of models had exceedance probability of 2%. Thus, it appears that expectation modulated both the feedforward and the feedback connections.

We also compared all nine individual models separately by examining their average posterior probabilities across subjects. The feedforward, feedback, and recurrent models in which the stimulus only influenced MT+ were generally preferred, with average posterior probabilities of 12%, 10%, and 44%, respectively. The models in which the stimulus directly influenced activity in both MT+ and DLPFC had average posterior probabilities of 5%, 2%, and 2% for the feedforward, feedback, and recurrent models, respectively. Lastly, models in which the stimulus modulated MT+, DLPFC, and the connectivity between MT+ and DLPFC had average posterior probabilities of 9%, 5%, and 10% for the feedforward, feedback, and recurrent models, respectively.

Finally, we looked at the strength of the connections between regions in the winning model (the recurrent model in which stimulus only modulates MT+; see Figure 6C). We were interested how expectation modulated the connections between bilateral MT and left DLPFC. Expectation significantly increased the strength of the forward connection from left MT to left DLPFC ($d = 0.098$; $t(20) = 3.43$, $p = 0.003$) and from right MT to left DLPFC ($d = 0.075$; $t(20) = 3.07$, $p = 0.006$). On the other hand, the feedback connections were negatively modulated by expectation though the effect was not significant (both $p$'s > .40).

**Discussion**

The likelihood of a perceptual event has a marked effect on our ability to perceive it (Bar, 2004). Here we examined how perceptual decisions about motion are influenced by inducing a prior expectation regarding the likelihood of motion direction. Behavioral results suggest that the prior cue changed subjects' expectation about upcoming motion direction, such that decisions were more accurate when subjects were provided with a valid cue (Figure 2). Moreover, dorsolateral prefrontal (DLPFC) cortex showed higher activity when subjects had a prior expectation about motion direction (Figure 3), and a modulation of activity as a function of whether the expectation was violated or not. Furthermore, between-subject activity differences in the the left DLPFC reflected individual differences in how much the expectation cue biased subjects' decisions. (Figure 4). Additionally, the left DLPFC showed enhanced recurrent connectivity with the motion sensitive area MT+ (which is likely representing the sensory evidence in this experiment), when subjects had an expectation about the direction of upcoming motion (Figure 5). Below we will interpret and discuss these results within the context of current models of perceptual decision making.

*Prior expectation increases activity in dorsolateral prefrontal cortex*

The prefrontal and parietal cortices are thought to play an important role in perceptual decision making (Gold and Shadlen, 2007; Heekeren et al., 2008; Kayser, Buchsbaum, et al., 2010; Kayser, Erickson, et al., 2010; Scheibe et al., 2010). Specifically, electrophysiological studies suggest that neural activity in these regions may reflect the accumulation of evidence during the decision

process (Shadlen and Newsome, 1996; Schall, 2003). When we compared trials in which subjects had a stimulus expectation with trials in which subjects had no stimulus expectation, we found heightened activity in both IPS and DLPFC. One interpretation of this finding is that the incorporation of prior expectation into the decision process changes the computations in these areas related to evidence accumulation.

However, it could be argued that the expectation cue simply induced a larger cognitive demand, because subjects needed to attend to both the cue and the stimulus for these type of trials. Although this interpretation can explain the fact that bilateral DLPFC and IPS are more active in the presence of expectation, it does not seem compatible with our exact pattern of results. In particular, the activation of DLPFC was additionally modulated by the *validity* of the cue: there was larger activity for invalidly cued than validly cued trials. This is consistent with a role of DLPFC in evidence accumulation, since subjects will need to accumulate more evidence for the correct motion direction when they have an invalid prior expectation. An alternative explanation for the activity difference in DLPFC between valid and invalid expectations could be that additional activity in the DLPFC arose during the invalidly-cued trials because of cognitive conflict (Ridderinkhof et al., 2004). However, this alternative explanation would not predict larger effective connectivity between sensory area MT+ and DLPFC when subjects had a prior expectation. This connectivity modulation is more in line with a specific role of the DLPFC in sending its predictions to sensory area MT+, as well as integrating the prior expectation and the sensory information provided by MT+. Also the fact that individual differences in criterion shift were correlated with the activity difference between predictive (valid and invalid) and non-predictive (neutral) cues in DLPFC is not in line with cognitive conflict. The activity in DLPFC is rather in line with earlier studies that observed activity modulations in DLPFC consistent with evidence accumulation (Kim and Shadlen, 1999; Heekeren et al., 2004; Heekeren et al., 2006; Ho et al., 2009), and therefore this region may play an important role in incorporating prior expectation into perceptual decisions.

Perhaps somewhat surprisingly, we did not find an overall increase in sensory area MT+ as a function of perceptual expectation, and the activity in MT+ only showed a weak negative correlation with the extent to which each subject used the predictive cues. Top-down modulations on sensory representations are well known (Gilbert and Sigman, 2007), and indeed previous studies have shown that expectation about motion direction can boost activity in direction-selective MT+ cells (Treue and Martínez Trujillo, 1999). In line with this, a previous neuroimaging study did find a modulation of MT+ activity when subjects had an expectation about upcoming motion (Shulman et al., 1999). One important difference between the study by Shulman et al. and the current experiment relates to detection vs discrimination. Shulman et al. asked subjects to detect motion, and subjects were given a cue that indicated the direction in which coherent motion could occur (up, down, left, or right). In that task, it is plausible that the activity of relevant MT+ neurons may be upregulated without suppressing other MT+ neurons. In our task, subjects were asked to discriminate between two motion directions (expanding or contracting). Thereby, the expectation cues may have led to simultaneous enhancement and suppression of two competing neuronal assemblies, precluding an overall activity increase. This interpretation could be directly tested in future research.

*Perceptual decisions and motor preparation*

In many previous studies, perceptual decisions are linked to specific motor responses throughout the experiment (Tosoni et al., 2008). This means that the accumulated evidence for the perceptual decision itself does not need to be held "online" but can be directly translated into the preparation of specific motor plans. This may explain why previous studies have reported results of perceptual decision-related activity in the motor cortex (Donner et al., 2009) or in the cortico-striatal network which is thought to be related to action selection (Forstmann et al., 2010). However, studies that dissociated the perceptual decision from the response modality have observed neural activity in parietal (Bennur and Gold, 2011) and prefrontal (Heekeren et al., 2006; Ferrera et al., 2009) cortex, regions that may be more related to the perceptual decision itself, rather than motor preparation. In our experiment, we have carefully minimized the possibility that our results would be contaminated by motor preparation, by informing subjects of the response mapping only after the presentation of the stimulus. Under these

circumstances, prior expectation modified the activity level in the DLPFC and connectivity between MT+ and DLPFC in a fashion that is in line with a role for this region in evidence accumulation. Although we did not find the same relationship in the parietal cortex, we did observe a general modulation of the presence vs. absence of expectation on neural activity in the anterior intraparietal sulcus. Therefore, it appears plausible that this region may also have a role in the accumulation of evidence in an effector-independent fashion (Bennur and Gold, 2011), in terms of jointly representing the prior expectation and the sensory evidence.

*Connectivity between MT+ and DLPFC*

Our PPI analysis demonstrated that expectation increased the effective connectivity between MT+ and DLPFC. Further, the DCM analysis suggested that this increased connectivity is likely driven by modulations of both the feedforward and the feedback connections. Increased feedback connectivity from DLPFC to MT+ could embody a change of gain setting in sensory regions by expectation (Shulman et al., 1999) and thus provide a mechanism of how prior information could change decision making, namely by frontal regions changing the responsiveness of relevant neurons in sensory cortex. Complementing this, increased feedforward connectivity could then be the result of this increased gain of the expected stimulus on prefrontal areas. This change of feedforward connectivity is consistent with previous neurophysiological research (Law and Gold, 2009).

*Relationship with other biasing effects on perceptual decision making*

Two ways that have been traditionally used to bias decision making is by either altering base rate of certain stimulus categories or by altering reward contingencies associated with certain stimulus categories. One may speculate that these top-down effects are hard to distinguish, since they have superficially similar effects on behavior (making a particular response more likely and its reaction time faster). Also, reward can change activity in both sensory (Serences, 2008) and decision-related (Rorie et al., 2010; Summerfield and Koechlin, 2010) areas. Conceptually, one can imagine that changing reward can influence behavior without changing subjects' belief as to what is actually the stimulus (e.g., "I do not really think the stimulus contains expanding motion, but since it will more

likely give me high reward, I am going to answer as such anyway"). Changing prior expectation, on the other hand, can plausibly change subject's beliefs and potentially the content of perception (Sterzer et al., 2008). Therefore, though both manipulations change behavior, the reward structure may dictate how behaviorally relevant the stimulus is, while the prior expectation pertains to its likelihood (Summerfield and Egner, 2009). These differences may explain why changing prior expectation may change the interaction between sensory and decision-making processes, while changing reward may selectively change the decision stage.

*Conclusion*

We provide behavioral and neural evidence for how prior expectation biases perceptual decisions. Behaviorally, prior expectations lead to both more biased and less sensitive perception. Neuroimaging data show that this bias is implemented by increasing activity in, as well as sensory drive to the dorsolateral prefrontal cortex.

**References**

Bar M (2004) Visual objects in context. Nat Rev Neurosci 5:617-629.

Bennur S, Gold JI (2011) Distinct Representations of a Perceptual Decision and the Associated Oculomotor Plan in the Monkey Lateral Intraparietal Area. J Neurosci 31: 913-921.

Brainard DH (1997) The Psychophysics Toolbox. Spatial Vision 10: 433-436.

Donner TH, Siegel M, Fries P, Engel AK (2009) Buildup of choice-predictive activity in human motor cortex during perceptual decision making. Current Biology: CB 19:1581-1585.

Egner T, Monti JM, Summerfield C (2010) Neural integration of top-down spatial and feature-based information in visual search. J Neurosci 30:16601-16608.

Enns JT, Lleras A (2008) What's next? New evidence for prediction in human vision. Trends Cogn Sci 12:327-333.

Esterman M, Yantis S (2009) Perceptual Expectation Evokes Category-Selective Cortical Activity. Cerebral Cortex (New York, NY: 1991).

Ferrera VP, Yanike M, Cassanello C (2009) Frontal eye field neurons signal changes in decision criteria. Nat Neurosci 12:1458-1462.

Forstmann BU, Brown S, Dutilh G, Neumann J, Wagenmakers EJ (2010) The neural substrate of prior information in perceptual decision making: a model-based analysis. Frontiers in human neuroscience 4:40.

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218-229.

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. Neuroimage 19:1273-1302.

Friston KJ, Holmes A, Poline JB, Price CJ, Frith CD (1996) Detecting Activations in PET and fMRI: Levels of Inference and Power. Neuroimage 4:223-235.

Gilbert CD, Sigman M (2007) Brain states: top-down influences in sensory processing. Neuron 54:677-696.

Gitelman DR, Penny WD, Ashburner J, Friston KJ (2003) Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. Neuroimage 19:200-207.

Gold JI, Shadlen MN (2007) The neural basis of decision making. Annual review of neuroscience 30:535-574.

Heekeren HR, Marrett S, Bandettini PA, Ungerleider LG (2004) A general mechanism for perceptual decision-making in the human brain. Nature 431:859-862.

Heekeren HR, Marrett S, Ruff DA, Bandettini PA, Ungerleider LG (2006) Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality. Proc Natl Acad Sci U S A 103:10023-10028.

Heekeren HR, Marrett S, Ungerleider LG (2008) The neural systems that mediate human perceptual decision making. Nat Rev Neurosci 9:467-479.

Ho TC, Brown S, Serences JT (2009) Domain general mechanisms of perceptual decision making in human cortex. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 29:8675-8687.

Kayser AS, Buchsbaum BR, Erickson DT, D'Esposito M. (2010) The functional anatomy of a perceptual decision in the human brain. J Neurophysiol 103:1179-1194.

Kayser AS, Erickson DT, Buchsbaum BR, D'Esposito M (2010) Neural representations of relevant and irrelevant features in perceptual decision making. J Neurosci 30:15778-15789.

Kim JN, Shadlen MN (1999) Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat Neurosci 2:176-185.

Law CT, Gold JI (2009) Reinforcement learning can account for associative and perceptual learning on a visual-decision task. Nat Neurosci 12:655-663.

Lund TE, Norgaard MD, Rostrup E, Rowe JB, Paulson OB (2005) Motion or activity: their role in intra- and inter-subject variation in fMRI. Neuroimage 26:960-964.

Macmillan N, Creelman C (2005) Detection theory: A user's guide. New York, NY: Cambridge University Press.

Noppeney U, Ostwald D, Werner S. (2010) Perceptual dicisions formed by accumulation of audiovisual evidence in prefrontal cortex. J Neurosci 30: 7434-7446.

Preuschhof C, Schubert T, Villringer A, Heekeren HR (2010) Prior information biases stimulus representations during vibrotactile decision making. J Cogn Neurosci 22:875-887.

Rajkowska G, Goldman-Rakic PS (1995) Cytoarchitectonic definition of prefrontal areas in the normal human cortex: II. Variability in locations of areas 9 and 46 and relationship to the Talairach Coordinate System. CerebCortex 5:323-337.

Ratcliff R, McKoon G (2008) The diffusion decision model: theory and data for two-choice decision tasks. Neural computation 20:873-922.

Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S (2004) The role of the medial frontal cortex in cognitive control. Science 306:443-447.

Rorie AE, Gao J, McClelland JL, Newsome WT (2010) Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. PLoS ONE, 5, e9308.

Schall JD (2003) Neural correlates of decision processes: neural and mental chronometry. Current opinion in neurobiology 13:182-186.

Scheibe C, Ullsperger M, Sommer W, Heekeren HR (2010) Effects of parametrical and trial-to-trial variation in prior probability processing revealed by simultaneous electroencephalogram/functional magnetic resonance imaging. J Neurosci 30:16709-16717.

Serences JT (2008) Value-based modulations in human visual cortex. Neuron 60:1169-1181.

Shadlen MN, Newsome WT (1996) Motion perception: seeing and deciding. Proc Natl Acad Sci U S A 93:628-633.

Shulman GL, Ollinger JM, Akbudak E, Conturo TE, Snyder AZ, Petersen SE, Corbetta M. (1999) Areas involved in encoding and applying directional expectations to moving objects. J Neurosci 19:9480-9496.

Stephan KE, Penny WD, Daunizeau J, Moran R, Friston KJ (2009) Bayesian model selection for group studies. NeuroImage 46: 1004-1017.

Sterzer P, Frith C, Petrovic P (2008) Believing is seeing: expectations alter visual awareness. Curr Biol 18:R697-698.

Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. Trends Cogn Sci.

Summerfield C, Egner T, Greene M, Koechlin E, Mangels J, Hirsch J (2006) Predictive codes for forthcoming perception in the frontal cortex. Science 314:1311-1314.

Summerfield C, Koechlin E (2008) A neural representation of prior information during perceptual inference. Neuron 59:336-347.

Summerfield C, Koechlin E (2010) Economic value biases uncertain perceptual choices in the parietal and prefrontal cortices. Front Hum Neurosci. 17:208.

Tosoni A, Galati G, Romani GL, Corbetta M (2008) Sensory-motor mechanisms in human parietal cortex underlie arbitrary visual decisions. Nature Neuroscience 11:1446-1453.

Treue S, Martínez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. Nature 399:575-579.

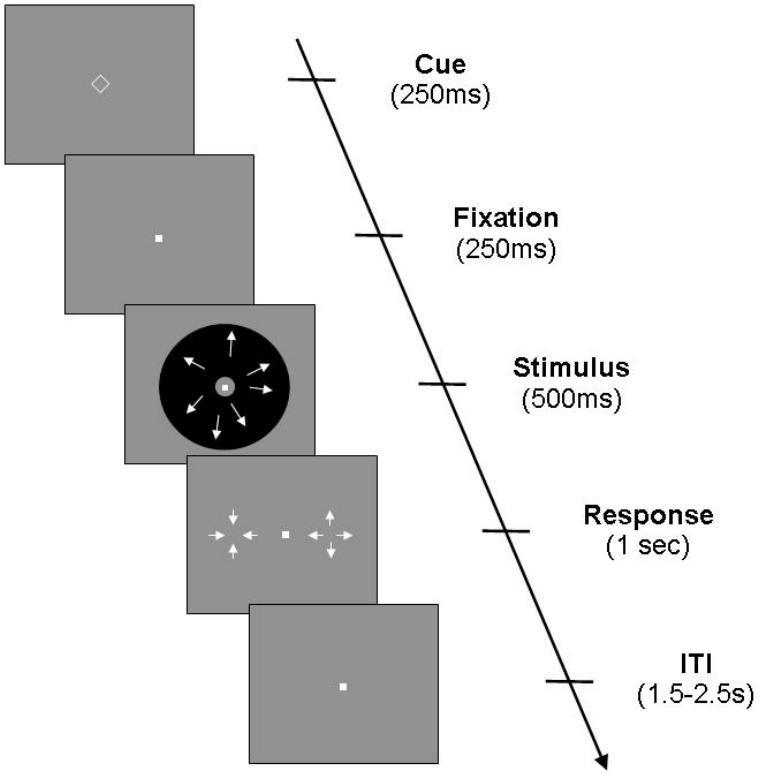


**Figure 1. Task design.** In each trial, subjects were asked to press keys to indicate whether a patch of dots was moving in an expanding or contracting direction. A cue in the form of a simple geometric shape indicated the likely direction of the motion. In half of the trials the cue was predictive of the motion direction (75% valid, 25% invalid), while in the other half of the trials a neutral cue was presented, which gave no information about the likely direction of the upcoming motion (i.e., a non-predictive cue). To perform optimally when the cue was predictive, subjects needed to combine the information from the cue and stimulus. The response mapping was only shown after the offset of the stimulus, so that subjects could not prepare for a motor response before the end of the motion presentation. Each trial lasted for between 3.5 and 4.5 seconds (chosen from a uniform distribution) and subjects had up to 1 second to give an answer after the offset of the stimulus.

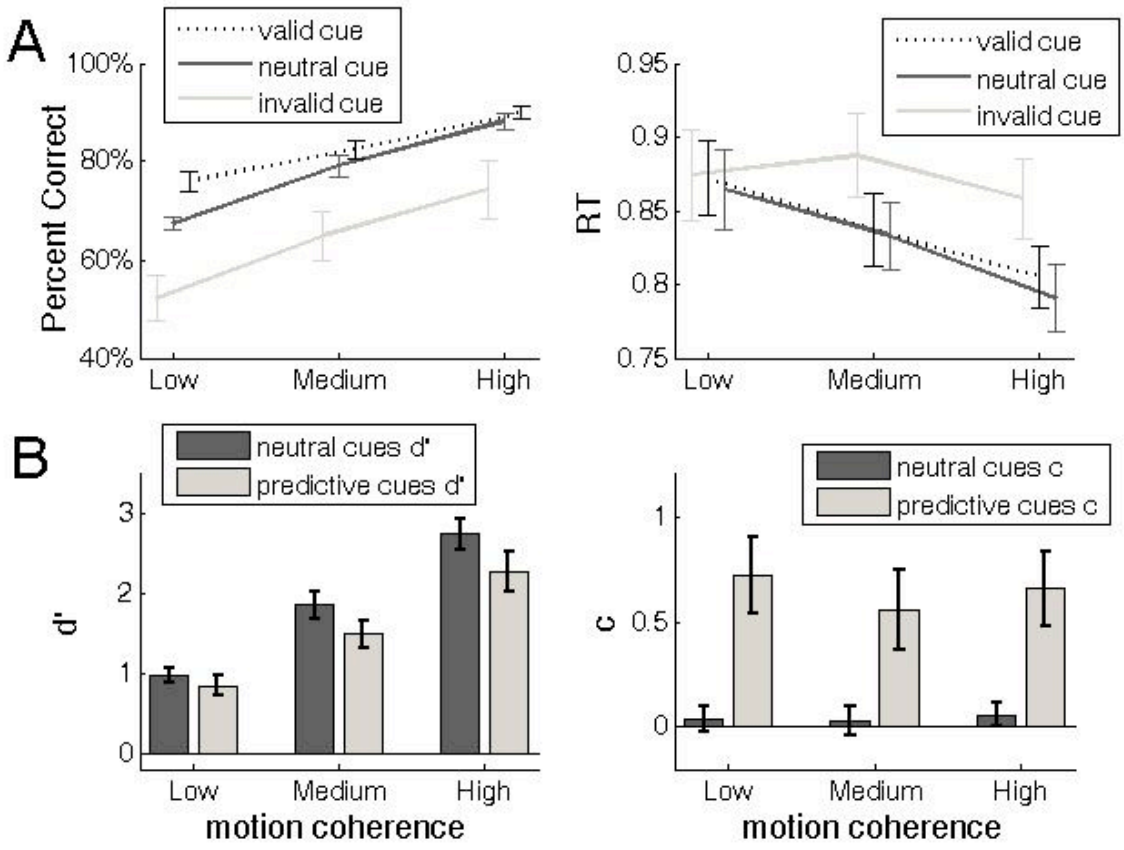


**Figure 2. Behavioral results.** A. Accuracy and reaction times are plotted as a function of motion coherence (low, medium, high) and cue type (valid, neutral, invalid). Overall, higher motion coherence led to higher accuracy and lower reaction times. Similarly, invalid cues decreased performance while valid cues improved it. Error bars show the standard error of the mean.

B. The signal detection theoretic (SDT) measures d' and c were computed independently for predictive (i.e., valid and invalid) and non-predictive (i.e., neutral) cues in order to gauge subjects' ability to do the task with/without the expectation provided by the cues. Surprisingly, d' was lower for predictive than for neutral cues. Subjects were unbiased (c ~ 0) when doing the task with neutral cues but shifted their criterion towards the expected percept when predictive cues were presented. Error bars show the standard error of the mean.

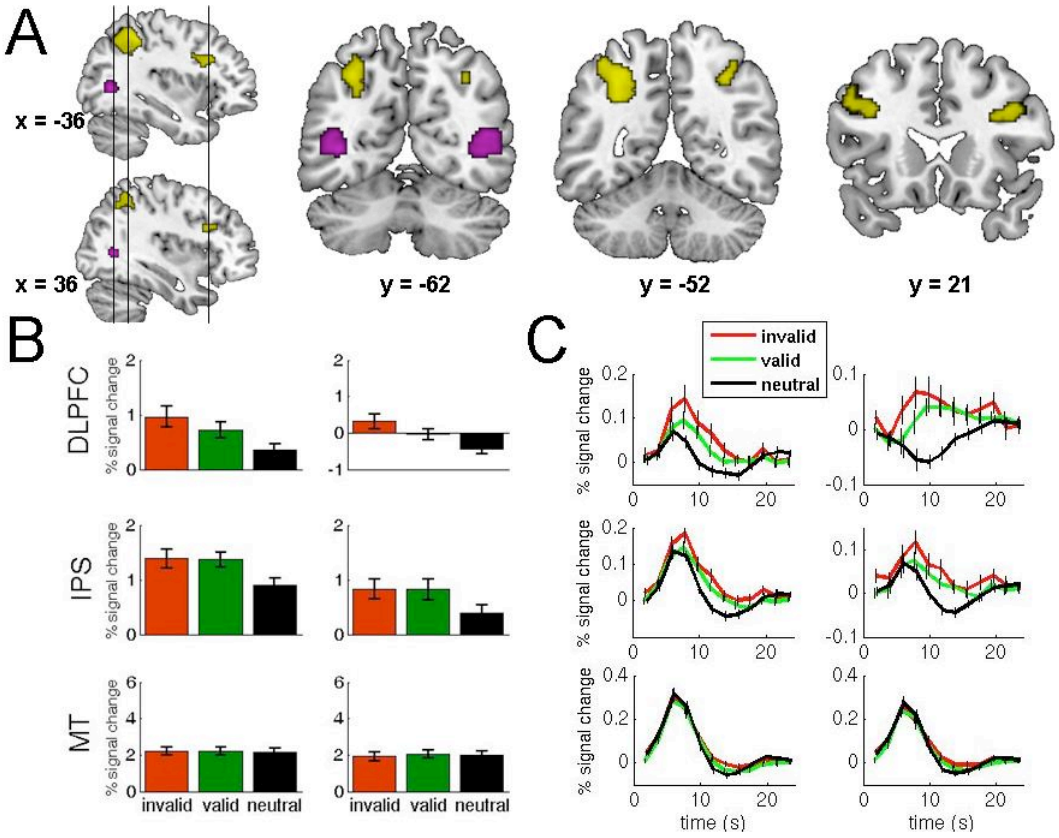


**Figure 3. Neural activation differences induced by expectation.** A. Larger activation was found bilaterally in both intraparietal sulcus (IPS) and bilateral dorsolateral prefrontal cortex (DLPFC) for trials in which subjects had an expectation than for trials in which subjects had no expectation (shown in yellow). Bilateral MT+ (shown in purple) was functionally localized using an independent localizer for each subject. B. Percent signal change is plotted for each of the three cue types (invalid, valid, and neutral) for left DLPFC, IPS and MT+ (left column) and its right hemisphere counterpart (right column). DLPFC showed larger activity for invalidly cued trials compared to validly cued trials (both $p$'s $< 0.02$). No such difference was found for IPS (both $p$'s $> 0.8$). There were no differences in MT+ for the differently cued trials (all $p$'s $> .2$). C. Timecourses for each of the six regions of interest are plotted for each trial type. Error bars represent the standard error of the mean.

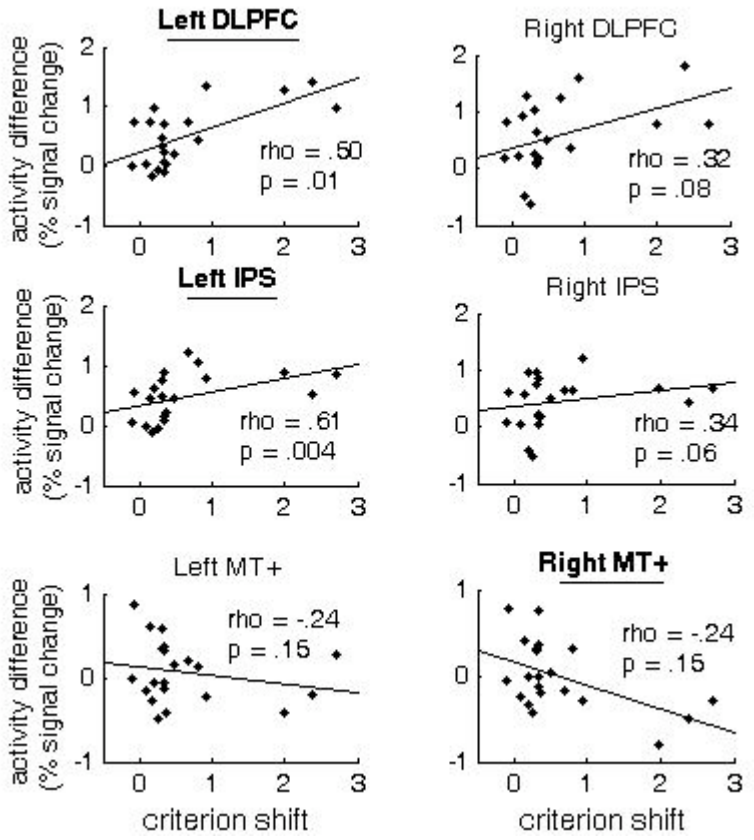


**Figure 4. Brain-behavior correlation of expectation-induced bias.** We observed a significant positive across-subjects correlation between the behavioral criterion shift induced by the prediction cue and the neural activity increase for predictived trials in left DLPFC (Spearman's rho = .50, $p$ = .01) and left IPS (Spearman's rho = .61, $p$ = .002). Only trends were found in right DLPFC and right IPS. On the other hand, this correlation was negative in left and right MT+ with the effect being significant in right MT+ (Spearman's rho = -.38, $p$ = .04). The activity increase was computed from the contrast predictive (valid and invalid) > non-predictive (neutral) trials. Criterion shift is a measure of the degree to which subjects adjusted their decision bias (see Methods for details). We used Spearman's rank correlation – a non-parametric test that is insensitive to extreme values in the variables. All significant correlations remain significant if Pearson's product-moment correlation was used.

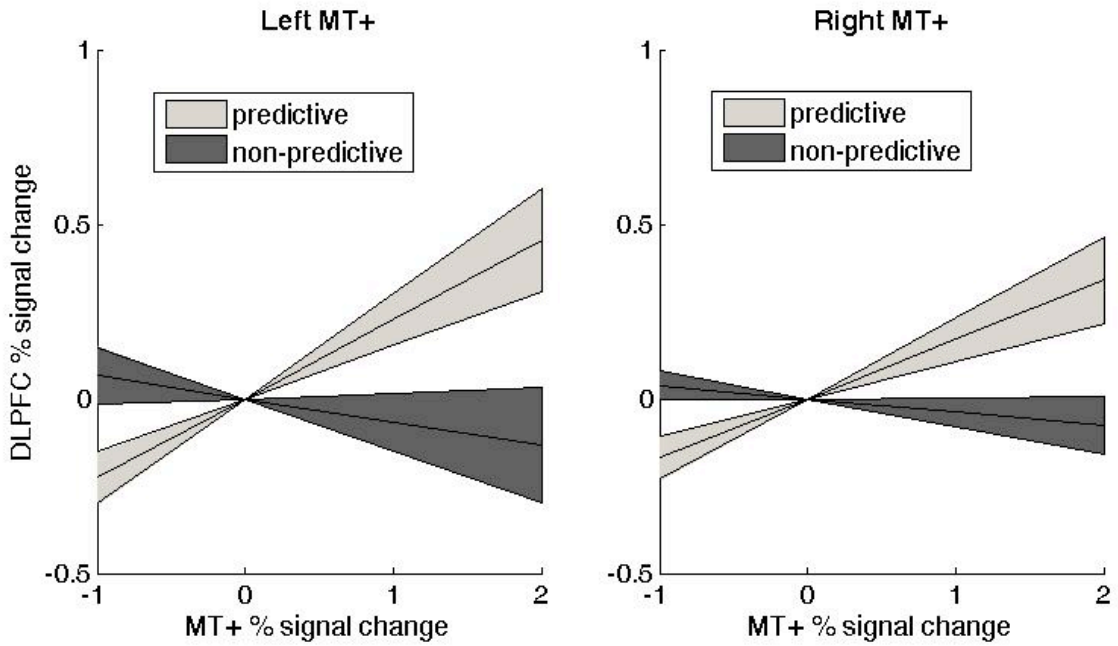


**Figure 5. Effective connectivity between bilateral MT+ and DLPFC.** We used a psychophysiological interaction (PPI) analysis to look at the effective connectivity between left DLPFC and left (left panel) and right (right panel) MT+. Left DLPFC was chosen because it was sensitive to the presence and validity of the cue (Figure 3), and correlated with the extent to which subjects shifted their criterion based on the predictive cues (Figure 4). We tested if the connectivity between MT+ and left DLPFC depended on the cue identity (computed from the contrast predictive [valid + invalid] > non-predictive [neutral]). Connectivity between MT+ and left DLPFC was higher when subjects had a prior expectation about upcoming motion direction, as revealed by the difference in the slope of regression between fMRI activity. The regression lines reflect averages across subjects. The shaded regions correspond to one SEM.

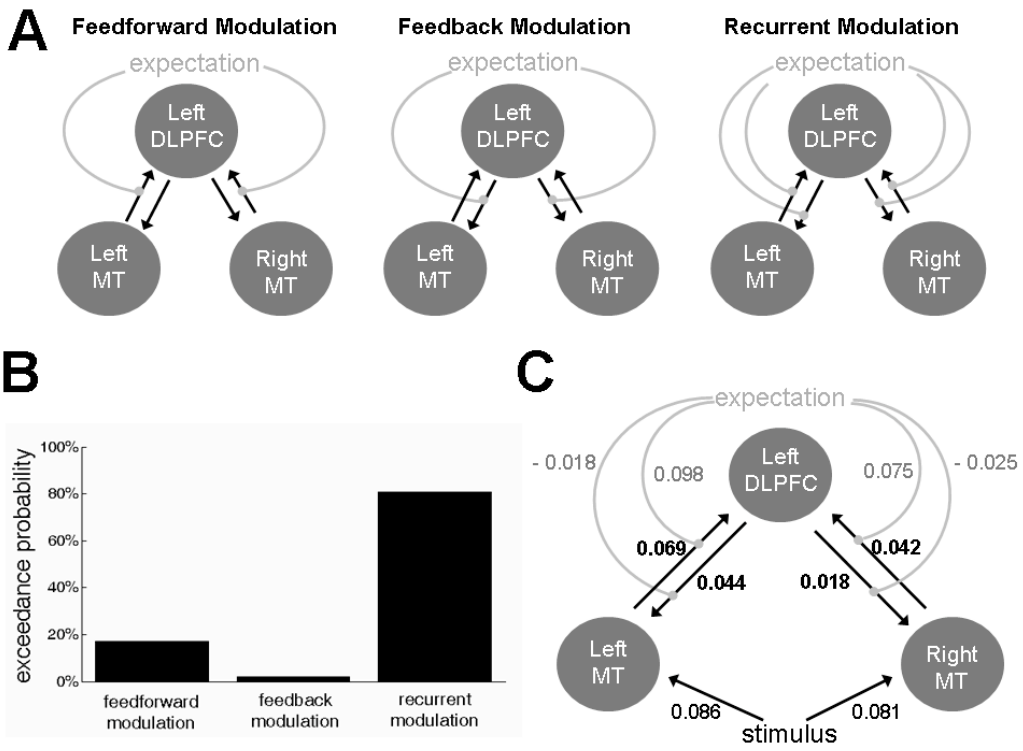


**Figure 6. Dynamic causal models testing the direction of connectivity between bilateral MT+ and left DLPFC.** A. Three families of models tested whether expectation modulated only the feedforward, the feedback, or both connections between MT+ and DLPFC. In all models, there were recurrent intrinsic connections between bilateral MT+ and left DLPFC. The contribution of the stimulus varied between models and is not depicted in this figure (see Methods for details). B. Bayesian model comparison was used to compute the exceedance probability for each of the three families of models. The recurrent family of models was the clear winner, indicating that expectation modulated both the feedforward and feedback connections between MT+ and DLPFC. C. All connections and their values are shown for the winning model (the recurrent model in which the stimulus modulated only MT+). Across subjects expectation significantly modulated the feedforward connection from left MT+ to left DLPFC ($p$ = .003) and from right MT+ to left DLPFC ($p$ = .006).

**Table 1.** Localization of activation difference for the contrast predictive (valid + invalid) > non-predictive (neutral)

| Anatomical region | T value | Cluster size | Corrected p value | Coordinates (x y z) |
|---|---|---|---|---|
| Left DLPFC | 4.49 | 257 | 0.001 | -50 22 36 |
| Right DLPFC | 4.27 | 162 | 0.01 | 40 20 30 |
| Left IPS | 5.68 | 1016 | < 0.001 | -40 -44 50 |
| Right IPS | 4.90 | 149 | 0.014 | 32 -56 44 |

Coordinates correspond to the standard Montreal Neurological Institute (MNI) brain. All results are cluster-level corrected for multiple comparisons.

# Appendix B

The influence of occipital TMS on brain processing

Dobromir A. Rahnev[1,2], Peter Kok[2], Moniek Munneke[2], Linda Bahdo[1], Floris P. de Lange[2], Hakwan Lau[1,2]

[1] Columbia University, Department of Psychology, New York, USA

[2] Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behavior, Netherlands

**Abstract**

Transcranial magnetic stimulation (TMS) has become one of the most important and widely used tools in cognitive neuroscience but we still have very limited understanding of the mechanisms through which TMS influences brain activity and cognitive processes. Here we investigated this issue by applying offline theta-burst transcranial magnetic stimulation (TBS) to the occipital cortex and investigated its influence in retinotopically-defined regions in early visual cortex using functional MRI. We found that occipital TBS reliably decreased the functional connectivity between regions of the early visual cortex. Such a decrease in functional connectivity could affect the strength of the perceptual signal, its variability, or both. We investigated this issue in a further experiment in which subjects performed a perceptual discrimination task and used wagering to indicate their level of certainty on each trial. The results showed that occipital TBS impaired subjects' performance by primarily on decreasing the strength of the perceptual signal with no reliable influence on its variability. We discuss the implications of these experiments for understanding the mechanisms of offline TMS on brain activity and perceptual processes, and note that these mechanisms may be different for online single-pulse TMS.

**Introduction**

Transcranial magnetic stimulation (TMS) has become one of the most important tools in cognitive neuroscience (Ziemann, 2011). TMS has been successfully applied to disturb visual processing (Amassian et al., 1989; Boyer et al., 2005; Corthout et al., 1999; Kammer, 2007; Kastner et al., 1998; Luber et al., 2007; Maccabee et al., 1991; Miller et al., 1996) and metacognition (Rounis et al., 2010), influence the speed of saccades (Bosch et al., 2012), impair working memory (Mottaghy et al., 2003), and even decrease people's self-control (Figner et al., 2010) and increase risk-taking (Knoch et al., 2006).

Nevertheless, despite the successful application of TMS to a number of domains and brain regions, our understanding of the mechanisms by which TMS affects brain processing is still lagging behind. Research in non-human animals has suggested that short TMS disrupts the temporal structure of activity by altering phase relationships between neural signals (Allen et al., 2007). Such results, coupled with a number of studies in humans that show behavioral decrements in tasks that depends on the stimulated region, has led to the "virtual lesion" view of TMS (Pascual-Leone et al., 2000), according to which the effects of brain stimulation are similar to temporarily lesioning the stimulated brain region. However, recently a number of studies have demonstrated that the effect of TMS extends beyond the focus of stimulation, often times to remote regions that are anatomically connected to the focal region (Fox et al., 2012; Garcia et al., 2011; Hampson and Hoffman, 2010; Ruff et al., 2009; Zanto et al., 2011).

Here we extended this line of work by investigating the influence of offline occipital TMS on the functional connectivity between retinotopically-defined regions in the visual cortex. To anticipate, we found that occipital TMS led to decreases in functional connectivity within retinotopically-defined regions in early visual cortex. A fellow-up experiment investigated

whether this effect on the brain's connectivity extended to behaviorally relevant outcomes. The results showed that offline TMS lowered performance by mainly influencing the strength of the perceptual signal rather than its variability.

## Methods

### *Subjects*

Five subjects participated in Experiment 1. A technical problem led to losing the functional MRI data from one subject in Experiment 1. The other four subjects in Experiment 1 were authors MM, PK, FdL, and LB (2 females, 23-32 years old).

Thirteen subjects participated in Experiment 2. Three subjects did not perceive phosphenes and were excluded from the analyses. One other subject did not come for all sessions of the experiment and was also excluded from the analyses. This left us with nine subjects (6 female, 19-26 years old).

Subjects from both experiments had normal or corrected-to-normal vision. They received detailed information about the potential side effects of TMS. A written informed consent was obtained from all subjects. The research was approved by the local ethics committee in which the experiment was performed (CMO region Arnhem-Nijmegen, The Netherlands).

### *TMS stimulation*

For both experiments, TMS was delivered with a Magstim Super Rapid Stimulator (Magstim Company Ltd, Whitland, UK), using a figure-of-eight coil with a diameter of 70 mm. We determined the phosphene threshold (PT) on a separate day before the fMRI scanning began using a procedure similar to Rahnev et al. (2012). Briefly, we first used a hunting procedure to

determine the optimal location for stimulation on the occipital cortex. The main axis of the coil was oriented parallel to the sagittal plane and the coil handle extended ventrally. We used suprathreshold intensity to find a location on the scull that produced a clear phosphene near the center of the visual field. We then proceeded to determine each individual's phosphene threshold in the following manner. Starting at 30% of the maximum stimulator output, we delivered single pulse TMS until we reached an intensity at which a subject reported five positive (phosphene present) before five negative (phosphene absent) responses; this intensity was chosen as the subject's "phosphene threshold." Theta-burst stimulation was delivered at 80% of the individual phosphene threshold. The resulting mean intensity of stimulation was 54.5% (SD = 12.8%) of maximum stimulator output for Experiment 1, and 46.6% (SD = 18.9%) of maximum stimulator output for Experiment 2. The intensity of TMS to the control sites (vertex in Experiment 1 and Pz in Experiment 2) was always the same as for the occipital cortex. No leg or other movement was elicited by vertex or Pz stimulation in any of the subjects.

*Experiment 1 procedure*

The experiment took place on two different days. On each day we collected functional magnetic resonance imagining (fMRI) of subjects' occipital cortex before and after they received theta burst transcranial magnetic stimulation (TBS). We stimulated the occipital cortex on one of the days, and the vertex of the head on the other day (the order of stimulation was counter-balanced between the subjects). The vertex was used as a control site. The phosphene threshold procedure described above was done on a different day prior to the first day of MRI acquisition.

After subjects arrived at the site, they were screened for metallic objects and were positioned in the MRI scanner. We then collected a retinotopy session using a rotating wedge for 8 minutes (314 scans), then a resting state scan for 10 minutes (399 scans), and then another identical rotating-wedge retinotopy session. Subjects were then escorted out of the scanner to an adjacent room with the transcranial magnetic stimulation (TMS) equipment. The TMS intensity was already pre-determined from a previous day (see above). In the vertex condition, we simply stimulated the site at the pre-determined intensity for 40 seconds during which we delivered five bursts of three 50-Hz pulses every second for a total of 600 pulses (Huang et al., 2005). In the occipital condition, we performed the same hunting procedure (see above) in order to find the to the optimal stimulation site and then applied TBS at the same pre-determined intensity. During the stimulation subjects were given a blindfold that they kept until they were placed back in the scanner. After the end of the stimulation, subjects were escorted back to the scanner as quickly as possible such that scanning resumed approximately five minutes after the stimulation. The same sequence of three scanning sessions followed.

In addition, for each subject, one of the two days ended with a few additional scans: an additional retinotopy session in which an annulus expanded and contracted for 10 minutes (410 scans), an anatomical scan of the visual cortex with the same coil as the one used to collect the functional data, and finally a whole brain anatomical scan with a separate coil.

*fMRI acquisition*

Images were acquired on a 3 tesla Trio MRI system (Siemens). Functional images were acquired using an 8-channel visual cortex coil, with a single-shot gradient echo-planar imaging sequence (repetition time, 1500 ms; echo time, 30 ms; 31 ascending slices; voxel size, 2 x 2 x 2 mm; flip angle, 70°; field of view, 220 mm). A high-resolution anatomical image

was acquired using a T1-weighted MPRAGE sequence (repetition time, 2300 ms; echo time, 3.52 ms; voxel size, 0.8 x 0.8 x 0.8 mm). In addition, another high-resolution anatomical image was acquired with a 32-channel coil using a T1-weighted MPRAGE sequence (repetition time, 2300 ms; echo time, 3.03 ms; voxel size, 1 x 1 x 1 mm).

*Retinotopy*

The boundaries of retinotopically organised areas in early visual cortex were identified using traveling-wave methods (Engel et al. 1994; Sereno et al. 1995). Visual field positions can be expressed as polar coordinates, i.e., in terms of angle and eccentricity. Angle was mapped by having subjects view a wedge, consisting of a flashing checkerboard pattern (3 Hz), first rotating clockwise for 9 cycles and then anticlockwise for another 9 cycles (at a rotation speed of 18 s/cycle). In a similar vein, eccentricity was mapped by presenting subjects with expanding (9 cycles, 18 s/cycle) and contracting (9 cycles) rings of flashing checkerboard patterns (3 Hz), centered on fixation. Fourier-based methods were used to obtain both the amplitude and the phase of the BOLD signal at the fundamental frequency of the stimuli (1/18 Hz). While the amplitude of the signal at this frequency, relative to the signal at other frequencies, can be seen an indication of the signal-to-noise ratio, the phase can be used to construct polar angle and eccentricity maps of the cortical surface. The borders of the visual areas (dorsal and ventral V1, V2, and V3) were defined on the basis of these maps, using Freesurfer (http://surfer.nmr.mgh.harvard.edu/). These retinotopic maps were then used to create regions of interest (ROIs) using MarsBaR (http://marsbar.sourceforge.net/).

*Analyses*

Once we identified the ROIs corresponding to different brain regions, we extracted timecourses for each of them for the resting state scans. For the main analysis we combined

the different sub-regions of V1, V2, and V3 so that we could look at the timecourse correlations for these regions as a whole. Besides that, we also performed planned tests between left and right V1, left and right V2, and left and right V3.

For each analysis we discarded the first six scans to allow for scanner equilibration. We further regressed out the timecourse of all cerebral spinal fluid voxels to control for non-neural influences. We also normalized each timecourse to have a mean of zero and a standard deviation of one. We then correlated the timecourses of different pairs of regions. Since the correlation coefficient r is not on a linear scale, we transformed each r values using the Fisher transformation (equivalent to the inverse hyperbolic tangent; Fisher, 1915):

$$r_{transformed} = \frac{1}{2} * (\ln(1+r) - \ln(1-r))$$

For each pair we obtained four different transformed correlation coefficients – pre occipital TBS ($r_{pre\_occ}$), post occipital TBS ($r_{post\_occ}$), pre vertex TBS ($r_{pre\_ver}$), and post vertex TBS ($r_{post\_ver}$). We were interested in the interaction of type of TBS and timing of stimulation (in order to be able to detect for a change in the correlations after occipital TBS). Therefore, we computed the interaction coefficient:

$$c_{int} = (r_{post\_occ} - r_{pre\_occ}) - (r_{post\_ver} - r_{pre\_ver})$$

Next, we created a confidence interval for the value of $c_{int}$ using bootstrapping. Specifically, for each pair of regions we considered the vectors of BOLD values for each session (pre occipital TBS, post occipital TBS, pre vertex TBS, and post vertex TBS). Then we shuffled the session labels of the pairs of BOLD values but kept the length of each session the same.

We then recomputed the interaction coefficient $c_{int}$ for the newly formed shuffled timecourses. By repeating this procedure 10,000 times, we computed the probability that $c_{int}$ would have a value as extreme as the one observed. Finally, we computed a fixed-effect confidence interval for the average group $c_{int}$ by running 10,000 simulations in which we took a randomly generated $c_{int}$ value from the shuffled data for each subject and averaged the four values.

*Experiment 2: Task and stimuli*

Experiment 2 took place over four different days. On the first day we determined subjects' phosphene thresholds as described above and then subjects practiced with the task. The next three sessions involved theta-burst stimulation to the occipital cortex, Pz (parietal midline; standard EEG nomenclature), and sham. As in Experiment 1, in the occipital stimulation session we determined the correct stimulation site by running an additional hunting procedure on the day of the theta-burst stimulation. For occipital and Pz stimulation, we used 80% of the originally determined phosphene threshold. For the sham stimulation we used 15% of the maximum stimulator output (so that the machine still produced the characteristic clicking sound) and placed the coil on top of the head but positioned it perpendicularly to the scalp so that only the side of the coil touched the head (and therefore only negligible amount of the magnetic field could reach the scull). The three TBS sessions were performed at least one week apart from each other.

Subjects' task was to indicate the tilt (clockwise or counterclockwise) of a grating presented at fixation. Each trial began with 50 ms presentation of the grating followed by a fixation period of 200 ms (Figure 3). On each trial the orientation of the grating was randomly determined to be titled 10º clockwise or 10º counterclockwise away from vertical. The grating was presented inside of an annulus (inner circle radius: 1.5º, outer circle radius: 4.5º). The

grating stimulus consisted of a noisy background composed of uniformly distributed intensity values (8% contrast) on top of which we added a grating (0.5 cycles/degree). Subjects were required to fixate on a small white square for the duration of the experiment. They were seated in a dim room about 50 cm away from a computer monitor. Stimuli were generated using Psychophysics Toolbox (Brainard, 1997) in MATLAB (MathWorks, Natick, MA) and were shown on a MacBook (13 inch monitor size, 1200 x 800 pixel resolution, 60 Hz refresh rate).

After each presentation subjects used one of four keys to give their response indicating the perceived orientation of the grating and a wager on whether they were correct. Subjects used the keys 1-4 indicating "certainly left", "guess left", "guess right", and "certainly right," respectively. A correct "certain" choice was awarded with two points while a correct "guess" choice was awarded with one point. An incorrect "guess" choice resulted in no points being won or lost but an incorrect "certain" choice resulted in a loss of two points. Thus, the optimal strategy was to choose the "certain" choice only when the probability of being correct exceeded 66.7%. We informed subjects of this contingency in order to guarantee that all subjects were aware of the optimal strategy and to try to ensure that subjects gave a sufficient number of both "guess" and "certain" responses. To further encourage optimal usage of the wagers, we gave the two subjects with highest final scores an additional cash prize.

Each trial lasted for exactly two seconds. Subjects had 1.8 seconds to give their response after the onset of the stimulus. Once a response was given, the text indicating the four possible answers (see Figure 3) disappeared and the next trial started exactly two seconds after the onset of the present trial. If a response was not given in the 1.8-second period subjects were penalized by a subtraction of four points and the text was removed at the end of the 1.8-

second period in order to avoid any potential interference with the processing of the stimulus in the next trial. Subjects completed five blocks of 140 trials each. Each block lasted 280 seconds (4 minutes and 40 seconds) and was followed by 20 seconds of rest for a total duration of five minutes per block. During the 20-second break after each block subjects were given feedback on their accumulated score from the last block.

Based on the results of the training session on day one, we chose a contrast for each subject that would produce ~80% correct responses. However, since some subjects tend to be biased and use conservative or liberal wagering strategies (Dienes and Seth, 2010), we decided to include two more levels of contrast: one that is 75% and the other 125% of the above contrast. The contrast level was chosen randomly on each trial and subjects were not explicitly informed about the presence of multiple contrast levels. Since contrast was not a variable of interest and was only used in order for subjects to use sufficiently both levels of wagering, for the purposes of the analyses we averaged across the three levels of contrast.

In the initial training session on day one, subjects practiced with the task over the course of five blocks of 120 trials each. In the first three blocks subjects received trial-by-trial feedback in order to learn to wager as optimally as possible. The last two blocks did not involve trial-by-trial feedback in order to prepare subjects for the actual experiment. In the blocks for which trial-by-trial feedback was provided, each trial was extended to 2.5 seconds in order for subjects to be able to see clearly the feedback. To provide further training with the task, each of the three TBS sessions started with 40 training trials that were not analyzed.

*Data analysis*

The signal detection theoretic measure d' was calculated by first coding each trial as a hit, miss, false alarm, or a correct rejection. Trials in which subjects reported that the stimulus was tilted in clockwise direction were coded as hits if the grating was indeed tilted clockwise and as false alarms otherwise. Trials in which subjects reported that the stimulus was tilted in counterclockwise direction were coded as correct rejections if the grating was indeed tilted in counterclockwise direction and as misses otherwise. Hit rate (HR) was computed as hits/(hits + misses) and false-alarm rate (FAR) was computed as false alarms/(false alarms + correct rejections). Finally, d' was calculated as:

$$d' = z(HR) - z(FAR)$$

where z is the inverse of the cumulative standard normal distribution that transforms HR and FAR into z scores.

We further estimated the signal detection theory (SDT) parameters related to the separation and variance of the internal distributions. To do so we made some standard assumptions: (1) the two stimuli used (clockwise and counterclockwise gratings) gave rise to Gaussian distributions on some decision axis, (2) the discrimination decision was made by comparing the signal on the decision axis with a criterion, while the wagering judgments were made by comparing the signal on the decision axis with two flanking criteria such that "certain" judgments were only produced in the two extremes of the decision axis, and (3) the criteria for perceptual decisions and wagering judgments were set in the same way for each day of the experiment (that is, they were the same for the pre-TMS and post-TMS sessions on each day). The last assumption is justified by our use of wagering judgments that encourage constant

placement of decision criteria, as well as by previous research that demonstrated that subjects used similar decision criteria even when that is clearly suboptimal (Gorea and Sagi, 2000).

Since we found no difference between Pz and sham stimulation (see Results section), we combined them as "control" conditions to be compared with occipital stimulation. For both occipital and control conditions we estimated the distance μ between the peaks of the Gaussian distributions and the standard deviation σ of these distributions. We did this separately for the pre-TMS and post-TMS sessions and compared the resulting values. In order to achieve a unique solution for μ and σ, we set the pre-TMS value of σ to an arbitrary constant value of 1 for both the occipital and control conditions and estimated the maximum likelihood values of the rest of the parameters. One subject used a very high percentage of high wagers making the estimation of μ and σ highly unstable; that subject was therefore excluded from this analysis. Finally, we used paired-sample t-test to compare the effect of occipital TMS on μ and σ, as well as other variables of interest such as d' and propensity for high wagering.

**Results**

We investigated whether occipital theta-burst transcranial magnetic stimulation (TBS) led to a significant change in the functional connectivity between regions from the early visual cortex during a resting state scan. To this end we defined V1, V2, and V3 (as well as their sub-regions) in each subject (Figure 1). We computed the functional connectivity for each pair of regions for each subject, and then transformed the correlation values using inverse hyperbolic tangent to account for the non-linear scale of correlation values (see Methods).

The results showed that occipital TBS led to a significant decrease in functional connectivity between V1 and V2, as well as between V1 and V3, for all four subjects (all $p$'s < .05; Figure 2). The decrease in functional connectivity between V2 and V3 after occipital TBS was significant for three out of the four subjects. In all three pairs, the effect was significant at the group level too. The effect of TBS was computed by subtracting the transformed correlation coefficients for the pre-TBS session from the post-TBS session (the absolute values of the transformed correlation coefficients were not meaningful). Significance was obtained using bootstrapping (see Methods).

We also checked the effect of occipital TBS on the functional connectivity between the left and right parts of V1, V2, and V3. For V1, we found that occipital TBS significantly decreased the functional connectivity of left and right V1 for all four subjects (all p's < .05). The same analysis showed a significant effect of occipital TBS on two of the four subjects for both V2 and V3. Nevertheless, the group effect was significant for all three regions.

Thus, Experiment 1 showed that occipital TBS caused a decrease in functional connectivity across the visual cortex. Nevertheless, at least two questions remained open. First, does this decrease in functional connectivity lead to behaviorally relevant outcomes? Second, would such effect on functional connectivity influence primarily the signal or the noise of the perceptual signal?

We investigated both of these questions in Experiment 2. We hypothesized that TBS to the occipital cortex would lead to a decrease in performance as compared to TBS to Pz (a control site) or sham TBS. We first checked for any differences between Pz and sham TBS. We compared the "TBS effect" for each of these two types of TBS by computing the change from

subjects' results before TBS administration (pre) to the results after TBS administration (post). We found no influence of TBS type on subjects' accuracy, willingness to use high wagers, overall points earned, or reaction times (all $p$'s > .2). We therefore averaged the Pz and sham TBS sessions as "control TBS" and compared that to occipital TBS.

Figure 4a plots the TBS effect for occipital and control TBS on the signal detection theoretic measure of capacity d' (see Methods). We found that occipital TBS significantly decreased d' ($t(8) = 2.94$, $p = .019$) while control TBS did not have a significant effect on it ($t(8) = -2.01$, $p = .079$). Critically, a paired-sample t-test demonstrated that the occipital TBS effect was significantly larger than the control TBS effect ($t(8) = 2.65$, $p = .029$).

Next we investigated subjects propensity to use high wagers (Figure 4b). We found that occipital and control TBS did not significantly modulate subjects' wagers ($p$'s > .1). However, a direct comparison between occipital and control TBS showed a trend for a lower propensity for high wagers after occipital TBS ($t(8) = 2.06$, $p = .073$), indicating that occipital TBS may have decreased not just subjects' objective ability to perform the task but also the certainty in their decisions.

These results extend our findings from Experiment 1 by confirming that occipital TBS had an effect on perception. In order to understand the nature of these effects, we analyzed the effect of TBS on the signal and noise of the perceptual signal. Since occipital TBS led to lower capacity d', we reasoned that this effect could be due either to signal decrease or noise increase. To distinguish between these two possibilities, we computed the distance between the signal detection distributions (that is, the signal $\mu$), as well as their variability (that is, the noise $\sigma$; see Methods). We found that, compared to control TBS, occipital TBS significantly

decreased the signal ($t(7) = 3.31$, $p = .013$; Figure 5). On the other hand, there was a trend for occipital TBS to also decrease the variability of the signal though this effect did not reach significance ($t(7) = 2.07$, $p = .078$). However, this effect on the variability of the signal was actually in the opposite direction of what one would expect if occipital TBS decreased d' by influencing the variance (i.e., if that were the case, the variance should have increased). Therefore, it appears that the main effect of occipital TBS in this experiment was to decrease the perceptual signal, which may have in turn led to a smaller or less reliable decrease in the variability of the signal.

**Discussion**

We investigated the influence of theta-burst transcranial magnetic stimulation (TBS) applied over the occipital cortex on the mechanisms of brain processing. In Experiment 1 we found that TBS decreased the functional connectivity across the early visual cortex. Experiment 2 extended these findings by demonstrating a behavioral effect on subjects' performance in a visual discrimination task. Further analyses demonstrated that TBS impacted the strength of the perceptual signal more than its variability. These findings shed new light on the mechanisms of TMS on neural processing.

Our results on the influence of TMS on the connectivity between separate brain regions extend previous research on the topic (Fox et al., 2012). For example, previous research has shown how phosphene perception can be affected by precisely timed conditioning pulses to motion sensitive area MT (Pascual-Leone and Walsh, 2001), parietal cortex (Silvanto et al., 2009), or frontal eye fields (FEF; Silvanto et al., 2006). Other research has reported remote effects of TMS outside of the visual cortex (Hampson and Hoffman, 2010; Ruff et al., 2009).

Nevertheless, to the best of our knowledge, no previous study has investigated the influence of TMS on the connectivity between parts of retinotopically-defined visual cortex.

The pattern of connectivity decreases suggested that TBS may have had a particularly strong influence on area V1. Indeed, V1 was the only area for which the connectivity between the left and right hemisphere parts was significantly decreased in all four subjects, while for V2 and V3 this effect was significant in only two out of the four subjects. This effect is somewhat surprising because even though it is commonly assumed that occipital TMS primarily targets the visual area V1 (Boyer et al., 2005; Laycock et al., 2007; Silvanto, Cowey, et al., 2005; Silvanto, Lavie, et al., 2005), newer research has shown that the site closest to the coil is likely to be dorsal V2 (Thielscher et al., 2010). In our study we did not use Brainsight or similar technology and thus cannot speculate about the precise area that was closest to the coil for each subject. Nevertheless, a recent study demonstrated that depending on each individual's occipital cortex folding, it could be possible to target V1 preferentially (Salminen-Vaparanta et al., 2012). Further, it is important to take into account that the proximity to the TMS coil is not the only factor that contributes to the effectiveness of TMS. It is well known that neurons are more excitable when the field is oriented along their axon and quite unexcitable to currents perpendicular to the axis (Rudin and Eisenman, 1954; Rushton, 1927). Consistent with this, some studies have found that the influence of TMS is greatest on regions whose cortical columns are longitudinally aligned to the direction of the current even if these regions are located sulcally while regions located in a gyrus – and thus physically closer to the coil – were relatively unaffected (Fox et al., 2004). Thus, it is possible that the columnar organization of V1 makes it more excitable to occipital TMS even if the coil is positioned closer to other areas of the visual cortex. In any case, since we stimulated the visual cortex offline, it is likely that persistent effects of stimulation were influenced by a

variety of additional factors including the pattern of connectivity of each region, the size of the regions, and the susceptibility of neurons to long term potentiation and depression.

It is currently unclear how exactly a decrease in functional connectivity (which we observed in Experiment 1) could lead to a signal loss (which we observed in Experiment 2). One possibility is that the signal present at earlier stages of the visual system does not propagate as well to downstream areas, and gradually loses part of its strength. Another possibility is that what is most affected is feedback from downstream areas that guide attention and set the perceptual template. More research is needed into the mechanisms by which functional connectivity affects psychological variables of interest.

Experiment 2 suggested that TBS may act by decreasing signal strength rather than affecting signal variability. However, recently we (Rahnev et al., 2012) and others (Ruzzoli et al., 2010; Schwarzkopf et al., 2011) have provided evidence that single pulse TMS may lead to an increase in perceptual strength rather than to signal loss (though see Harris et al., 2008; Ruzzoli et al., 2011 for studies that reached the opposite conclusion). Thus, it is important to note that the online and offline TMS are likely to affect neurons in a different fashion. While online TMS induces firing in a subset of the neurons in the stimulated area that interferes with signal processing (Allen et al., 2007), offline TMS leads to a lowered excitability of the stimulated area (Allen et al., 2007; Huang et al., 2005). Thus, it is likely that the precise nature of impairments also differs substantially between online and offline TMS. Finally, it is also possible that TMS stimulation interacts with the specific task at hand, which would explain the heterogeneity of findings in the single-pulse TMS literature.

Our results are in line with previous research that shows that TBS to the occipital cortex leads to increased phosphene thresholds (Franca et al., 2006). However, there is less consistency among previous studies on the effects of offline TMS on visual acuity: while some studies found decreased performance (Antal et al., 2002; Kosslyn et al., 1999), others reported increased visual acuity (Thompson et al., 2008; Waterston and Pack, 2010). There were large differences in the frequency, intensity, and location of the stimulation, the size and location of the stimuli in the visual field, and the nature of the task. All of these make the comparison between the current study and previous research difficult. Our results from Experiment 1 suggest that one of the effects of offline TMS is to decrease the functional connectivity between regions of the visual cortex. However, it is conceivable that even though many tasks would suffer from such connectivity decrease, other tasks may actually benefit. Details like the eccentricity of the stimulus in the visual field, or whether subjects are asked to perform detection versus discrimination tasks, may lead to different behavioral outcomes. More systematic research is needed in order to map out the influences of various factors such as stimulation intensity, location of the stimuli on the screen, and precise nature of the task on the effects of TMS stimulation.

One important caveat of our work is that we only applied TBS when subjects were at rest with their eyes closed. However, previous research has demonstrated that the level of activation during stimulation is an important factor for the behavioral effect of theta-burst stimulation (Silvanto et al., 2007). This research has suggested that TBS has the highest impact on the least active neural population. Thus, it remains an open question what pattern of connectivity decreases or increases would TBS produce if one subpopulation of neurons is active during theta-burst stimulation.

In conclusion, we found that theta-burst transcranial magnetic stimulation delivered at rest caused a decrease in connectivity among regions in the early visual cortex. This connectivity decrease led to a decrease in the signal – but not an increase in the noise – in a discrimination task. Therefore, the effects of offline TBS extend beyond the site of stimulation and impact the brain wiring.

**References**

Abrahamyan, A., Clifford, C. W., Arabzadeh, E., & Harris, J. A. (2011). Improving visual sensitivity with subthreshold transcranial magnetic stimulation. *J Neurosci, 31*(9), 3290-3294.

Allen, E.A., Pasley, B.N., Duong, T., & Freeman, R.D. (2007). Transcranial Magnetic Stimulation Elicits Coupled Neural and Hemodynamic Consequences. *Science, 317*, 1918-1921.

Amassian, V. E., Cracco, R. Q., Maccabee, P. J., Cracco, J. B., Rudell, A., & Eberle, L. (1989). Suppression of visual perception by magnetic coil stimulation of human occipital cortex. *Electroencephalography and Clinical Neurophysiology, 74*, 458-462.

Antal ACA, Kincses TZ, Nitsche MA, Bartfai O, Demmer I, Sommer M, Paulus W. (2002). Pulse configuration-dependent effects of repetitive transcranial magnetic stimulation on visual perception. *Neuroreport, 13*, 2229–2223.

Bosch SE, Neggers SF, Van der Stigchel S. (2012). The Role of the Frontal Eye Fields in Oculomotor Competition: Image-Guided TMS Enhances Contralateral Target Selection. *Cerebral Cortex* [Epub ahead of print].

Boyer, J.L., Harrison, S., Ro, T. (2005). Unconscious processing of orientation and color without primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 16875-16879.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433-436.

Breitmeyer, B.G., Ro, T., & Ogmen, H. (2004). A comparison of masking by visual and transcranial magnetic stimulation: Implications for the study of conscious and unconscious processing. *Consciousness and Cognition, 13*, 829-843.

Corthout, E., Uttl, B., Walsh, V., Hallett, M. & Cowey, A. (1999). Timing of activity in early visual cortex as revealed by transcranial magnetic stimulation. *NeuroReport, 10*, 2631–2634.

Dienes, Z., & Seth, A. (2010). Gambling on the unconscious: a comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Consciousness and Cognition, 19*, 674-681.

Figner, B., Knoch, D., Johnson, E.J., Krosch, A.R., Lisanby, S.H., Fehr, E., & Weber, E.U. (2010). Lateral prefrontal cortex and self-control in intertemporal choice. *Nature Neuroscience, 13*, 538–539.

Fisher, R.A. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika, 10*, 507-521.

Fox MD, Halko MA, Eldaief MC, Pascual-Leone A. (2012). Measuring and manipulating brain connectivity with resting state functional connectivity magnetic resonance imaging (fcMRI) and transcranial magnetic stimulation (TMS). *Neuroimage* [Epub ahead of print].

Fox PT, Narayana S, Tandon N, Sandoval H, Fox SP, Kochunov P, Lancaster JL. (2004). Column-based model of electric field excitation of cerebral cortex. *Human Brain Mapping, 22*, 1-14.

Franca M, Koch G, Mochizuki H, Huang Y-Z, Rothwell JC (2006) Effects of theta burst stimulation protocols on phosphene threshold. *Clinical Neurophysiology, 117*, 1808–1813.

Garcia JO, Grossman ED, Srinivasan R. (2011). Evoked potentials in large-scale cortical networks elicited by TMS of the visual cortex. *Journal of Neurophysiology, 106*, 1734-1746.

Gorea, A. & Sagi, D. (2000). Failure to handle more than one internal representation in visual detection tasks. *Proceedings of the National Academy of Sciences of the United States of America, 97*(22), 12380-12384.

Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics*. New York: John Wiley & Sons, Inc.

Hampson M, Hoffman RE. (2010). Transcranial magnetic stimulation and connectivity mapping: tools for studying the neural bases of brain disorders. *Frontiers of System Neuroscience, 4*, 40.

Harris, J. A., Clifford, C. W., & Miniussi, C. (2008). The functional effect of transcranial magnetic stimulation: signal suppression or neural noise generation? *Journal of Cognitive Neuroscience, 20*(4), 734-740.

Huang Y-Z, Edwards MJ, Rounis E, Bhatia KP, Rothwell JC (2005). Theta Burst Stimulation of the Human Motor Cortex. *Neuron, 45*, 201-206.

Jolij, J., & Lamme, V. A. F. (2005). Repression of unconscious information by conscious processing: evidence from affective blindsight induced by transcranial magnetic stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(30), 10747-51.

Kammer, T. (2007). Masking visual stimuli by transcranial magnetic stimulation. *Psychological Research, 71*, 659-666.

Kastner, S., Demmer, I., & Ziemann, U. (1998). Transient visual field defects induced by transcranial magnetic stimulation over human occiptal lobe. *Experimental brain research, 118*, 19-26.

Knoch D, Gianotti LR, Pascual-Leone A, Treyer V, Regard M, Hohmann M, Brugger P. (2006). Disruption of right prefrontal cortex by low-frequency repetitive transcranial

magnetic stimulation induces risk-taking behavior. *Journal of Neuroscience, 2*6, 6469-6472.

Koivitso, M., Mantyla, T., & Silvanto, J. (2010). The role of early visual cortex (V1/V2) in conscious and unconscious visual perception. *Neuroimage, 51*, 828-834.

Koivitso, M., Railo, H., & Salminen-Vaparanta, N. (2011). Transcranial magnetic stimulation of early visual cortex interferes with subjective visual awareness and objective forced-choice performance. *Consciousness and Cognition, 20*(2), 288-298.

Kosslyn SM, Pascual-Leone A, Felician O, Camposano S, Keenan JP, Thompson WL, Ganis G, Sukel KE, Alpert NM. (1999) The Role of Area 17 in Visual Imagery: Convergent Evidence from PET and rTMS. *Science, 284*, 167–170.

Laycock, R., Crewther, D.P., Fitzgerald, P.B., & Crewther, S.G. (2007) Evidence for fast signals and later processing in human V1/V2 and V5/MT+: A TMS study of motion perception. *Journal of Neurophysiology, 98*, 1253-1262.

Luber, B., Stanford, A., Malaspina, D., & Lisanby, S.H. (2007) Revisiting the Backward Masking Deficit in Schizophrenia: Individual Differences in Performance and Modeling With Transcranial Magnetic Stimulation. *Biological Psychiatry, 62*, 793-799.

Maccabee, P.J., Amassian, V.E., Cracco, R.Q., Cracco, J.B., Rudell, A.P., Eberle, L.P., & Zemon, V. (1991). Magnetic coil stimulation of human visual cortex: studies of perception. *Electroencephelogr Clin Neurophysiol 43*, 111-120.

Miller, B.M., Fendrich, R., Eliassen, J.C., Demirel, S., & Gazzaniga, M.S. (1996). Transcranial magnetic stimulation: delays in visual suppression due to luminance changes. *Neuroreport, 7*, 1740-1744.

Mottaghy FM, Pascual-Leone A, Kemna LJ, Töpper R, Herzog H, Müller-Gärtner HW, Krause BJ. (2003). Modulation of a brain-behavior relationship in verbal working memory by rTMS. *Brain Res Cogn Brain Res, 15*, 241-249.

Overgaard, M., Nielsen, J.F., & Fuglsang-Frederiksen, A. (2004). A TMS study of the ventral projections from V1 with implications for the finding of neural correlates of consciousness. *Brain and Cognition, 54*, 58-64.

Pascual-Leone, A., & Walsh, V. (2001). Fast back projections from the motion to the primary visual area necessary for visual awareness. *Science, 292*, 510-512.

Pascual-Leone, A., Walsh, V., & Rothwell, J. (2000). Transcranial magnetic stimulation in cognitive neuroscience–virtual lesion, chronometry, and functional connectivity. *Current Opinion in Neurobiology, 10*, 232–237.

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F.P., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience, 14*, 1513-1515.

Rahnev, D.A., Maniscalco, B., Luber, B., Lau, H., & Lisanby, S.H. (2012). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *Journal of Neurophysiology, 107*, 1556-1563.

Ro, T., Breitmeyer, B., Burton, P., Singhal, N. S., & Lane, D. (2003). Feedback contributions to visual awareness in human occipital cortex. *Curr Biol, 13*(12), 1038-1041.

Ro, T., Shelton, D. J. M., Lee, O. L., & Chang, E. (2004). Extrageniculate mediation of unconscious vision in transcranial magnetic stimulation induced blindsight. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 9933-9935.

Rounis E, Maniscalco B, Rothwell JC, Passingham RE, & Lau H (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience, 1*, 1-11.

Rudin, D.O., & Eisenman, G. (1954). The action potential of spinal axon in vitro. *Journal of General Physiology, 37*, 505–538.

Ruff, C.C., Driver, J., Bestmann, S. (2009). Combining TMS and fMRI: from 'virtual lesions' to functional-network accounts of cognition. *Cortex, 45*, 1043-1049.

Rushton, W.A. (1927). Effect upon the threshold for nervous excitation of the length of nerve exposed and the angle between current and nerve. *Journal of Physiology, 63*, 357–377.

Ruzzoli, M., Abrahamyan, A., Clifford, C. W. G., Marzi, C. A., Miniussi, C., and Harris, J. A. (2011). The effects of TMS on visual motion sensitivity: an increase in neural noise or a decrease in signal strength? *J Neurophysiol, 106*, 138-143.

Ruzzoli, M., Marzi, C. A., & Miniussi, C. (2010). The neural mechanisms of the effects of transcranial magnetic stimulation on perception. *J Neurophysiol, 103*(6), 2982-2989.

Salminen-Vaparanta, N., Noreika, V., Revonsuo, A., Koivisto, M., & Vanni, S. (2012). Is selective primary visual cortex stimulation achievable with TMS? *Human Brain Mapping, 33*, 652-665.

Schwarzkopf, D. S., Silvanto, J., & Rees, G. (2011). Stochastic resonance effects reveal the neural mechanisms of transcranial magnetic stimulation. *J Neurosci., 31*(9), 3143-3147.

Siebner, H. R., Hartwigsen, G., Kassuba, T., & Rothwell, J. C. (2009). How does transcranial magnetic stimulation modify neuronal activity in the brain? Implications for studies of cognition. *Cortex, 45*(9), 1035-1042.

Silvanto, J., Cowey, A., Lavie, N., & Walsh, V. (2005). Striate cortex (V1) activity gates awareness of motion. *Nature Neuroscience, 8*, 143-144.

Silvanto, J., Lavie, N., & Walsh, V. (2005). Double dissociation of V1 and V5/MT activity in visual awareness. *Cerebral cortex, 15*, 1736-1741.

Silvanto, J., Lavie, N., Walsh, V. (2006). Stimulation of the human frontal eye fields modulates sensitivity of extrastriate visual cortex. *Journal of Neurophysiology, 96*, 941–945.

Silvanto. J, & Muggleton, N.G. (2008). New light through old windows: moving beyond the "virtual lesion" approach to transcranial magnetic stimulation. *Neuroimage, 39*, 549-552.

Silvanto, J., Muggleton, N.G., Cowey, A., & Walsh, V. (2007). Neural activation state determines behavioral susceptibility to modified theta burst transcranial magnetic stimulation. *European Journal of Neuroscience, 26*, 523-528.

Silvanto, J., Muggleton, N., Lavie, N., & Walsh, V. (2009). The perceptual and functional consequences of parietal top-down modulation on the visual cortex. *Cerebral Cortex, 19*, 327-330.

Thielscher, A., Reichenbach, A., Uğurbil, K., & Uludağ, K. (2010). The cortical site of visual suppression by transcranial magnetic stimulation. *Cerebral Cortex, 20*(2), 328-338.

Thompson B, Mansouri B, Koski L, Hess RF (2008) Brain Plasticity in the Adult: Modulation of Function in Amblyopia with rTMS. *Current Biology, 18*, 1067–1071.

Waterston ML, Pack CC (2010). Improved discrimination of visual stimuli following repetitive transcranial magnetic stimulation. *PLoS One, 5*, e10354.

Zanto TP, Rubens MT, Thangavel A, Gazzaley A. (2011). Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nature Neuroscience, 14*, 656-661.

Ziemann, U. (2011). Transcranial Magnetic Stimulation at the Interface with Other Techniques : A Powerful Tool for Studying the Human Cortex. *The Neuroscientist, 17*, 368-381.

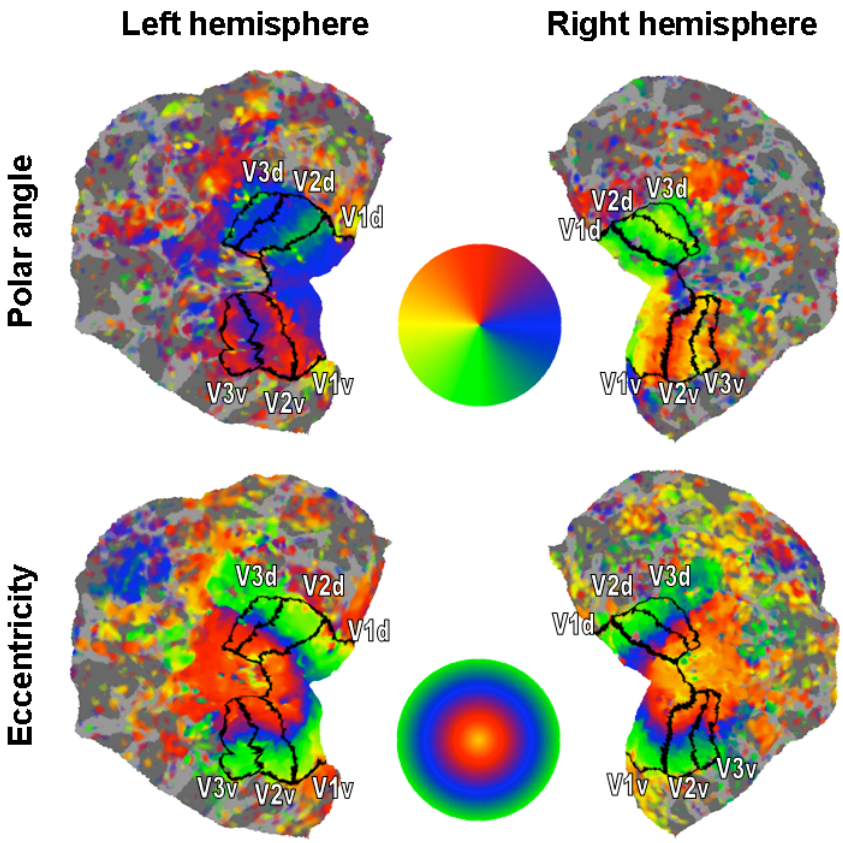


**Figure 1. Retinotopically-defined regions in early visual cortex.** The maps for polar angle (upper part) and eccentricity (lower part) are shown for both the left hemisphere (left part) and right hemisphere (right part) and. The maps are overlaid on the flattened occipital cortex of a representative subject. Black lines represent borders of the early visual areas V1, V2, and V3. v = ventral, d = dorsal.

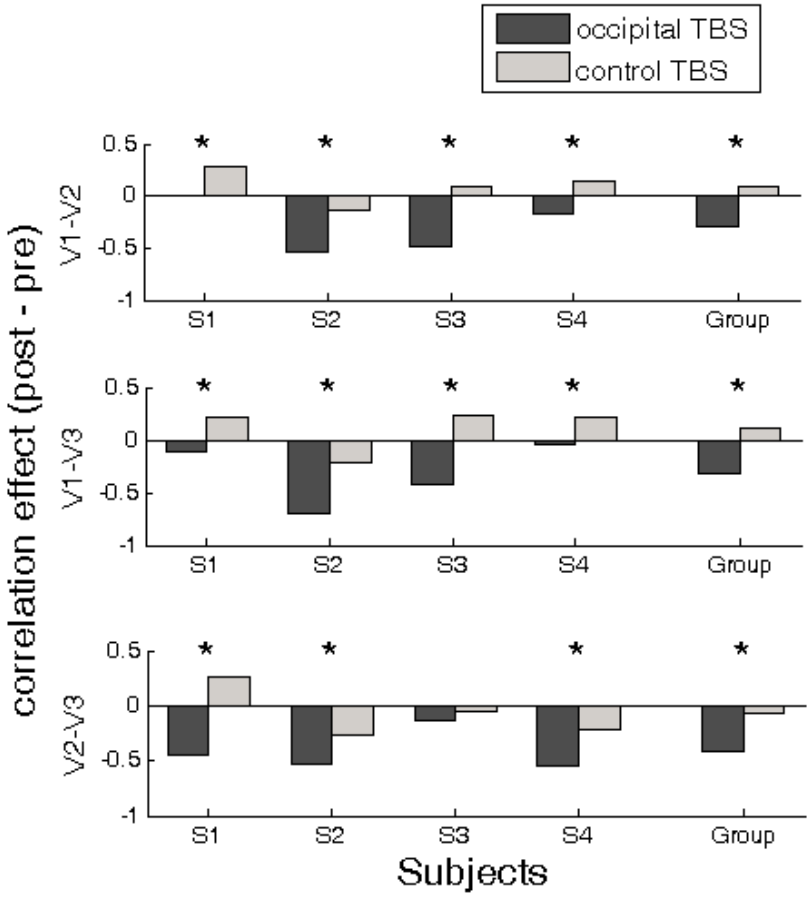


**Figure 2. Effect of TBS on the functional connectivity between regions in the early visual cortex.** Compared to control (vertex) theta-burst stimulation, occipital TBS led to a significant decrease in resting state functional connectivity between V1 and V2 for all four subjects, i.e., the difference in the strength of the correlation before and after TBS was greater for occipital than for control TBS. The same was true for the connectivity between V1 and V3, while the connectivity between V2 and V3 was significantly impaired for three out of the four subjects. A fixed effects analysis showed that all effects were significant at the group level. Significance values were determined using bootstrapping. * $p < .05$

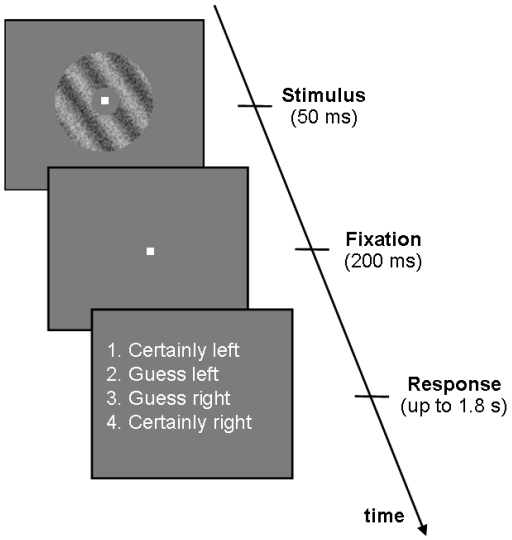


**Figure 3. Experiment 2 task.** Subjects discriminated whether a grating was tilted 10º to the right (clockwise) or left (counterclockwise) from vertical. Subjects also wagered on their level of certainty such that a high confidence response earned them more points when they were correct but made them lose more points when they were wrong (see Methods for details). Each trial lasted exactly two seconds and missed responses were punished with losing points.

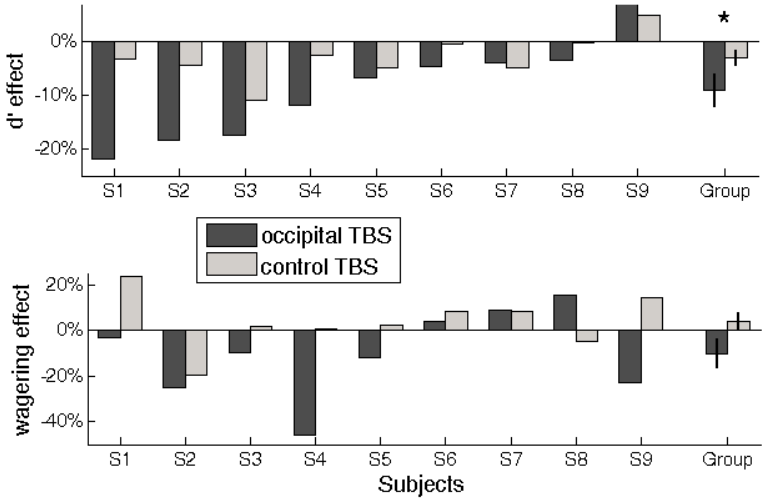


**Figure 4. Effect of TBS on capacity d' and propensity for high wagers.** Occipital TBS led to a significantly higher decrease in d' compared to control TBS ($p$ = .029). There was also a trend for occipital TBS to lower subjects' propensity to use high wagers though that effect did not reach significance ($p$ = .073). Error bars represent SEM. * $p < .05$

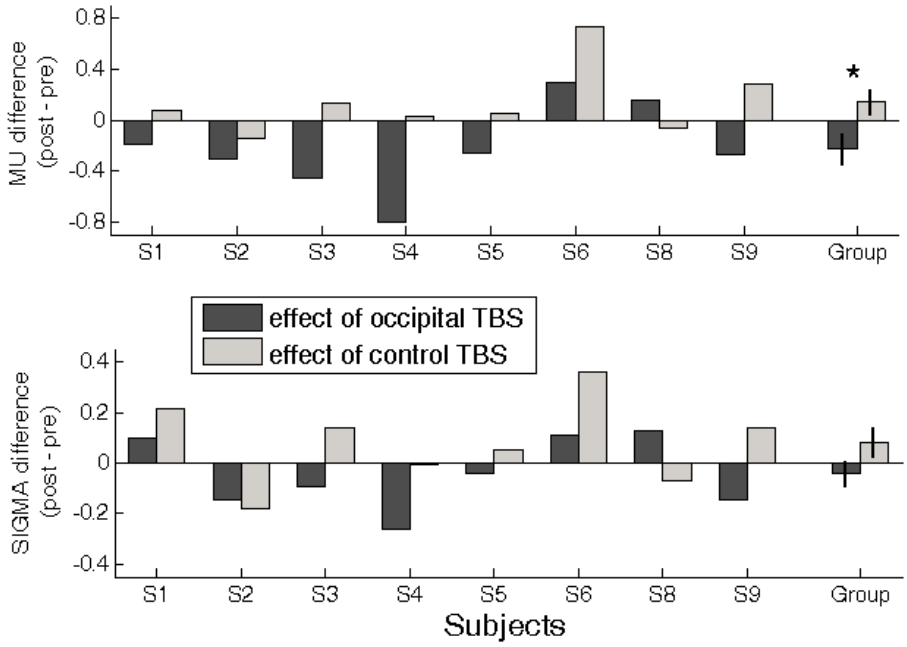


**Figure 5. Effect of TBS on the perceptual signal (μ) and its variability (σ).** We estimated the effect of TBS on the separation and variability of the internal distributions. Since subject S7 had very high number of high confidence wagers the estimated values of μ and σ were highly variable and the subject was excluded from this analysis. Overall, compared to control theta-burst stimulation, occipital TBS decreased the strength of the signal (μ; $p = .013$) without significantly affecting its variability ($p = .078$). Thus the behavioral impairments reported in Figure 4 appeared to be due to a signal loss rather than noise increase. Error bars represent SEM. * $p < .05$